

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 18:02:09 ; Search time 55.1415 Seconds

(without alignments)  
4469.091 Million cell updates/sec

Title: US-09-823-394-2

Sequence: 1 MKTFSFFLSTVTLFFFSF.....GFTLEMDMSIKVEPGL 1196

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

7. number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP\_ARCHA:.\*  
2: SP\_BACTERIA:.\*  
3: SP\_FUNGI:.\*  
4: SP\_HUMAN:.\*  
5: SP\_INVERTEBRATE:.\*  
6: SP\_MAMMAL:.\*  
7: SP\_MHC:.\*  
8: SP\_ORGANELLE:.\*  
9: SP\_PHAGE:.\*  
10: SP\_PLANT:.\*  
11: SP\_PODENT:.\*  
12: SP\_VIRUS:.\*  
13: SP\_VERTEBRATE:.\*  
14: SP\_UNCLASSIFIED:.\*  
15: SP\_VIRUS:.\*  
16: SP\_BACTERIAP:.\*  
17: SP\_ARCHAEP:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6140	99.3	1196	10	O22476
2	3050.5	49.3	1121	10	O242F3
3	2643.5	42.8	1192	10	O242F3
4	2633.5	42.6	1164	10	O242F3
5	2623.5	42.4	1166	10	O242F3
6	2582.5	41.8	1166	10	O242F3
7	2509.5	40.6	1143	10	O242F3
8	2240	36.2	1110	10	O242F3
9	1608.5	26.0	1110	10	O242F3
10	1593.5	25.8	1129	10	O242F3
11	1503	24.3	1120	10	O242F3
12	1419	23.0	1132	10	O242F3
13	1402.5	22.7	1106	10	O242F3
14	1380.5	22.3	1102	10	O242F3
15	1380	22.3	1079	10	O242F3
16	1370.5	22.2	1141	10	O242F3

17	1367	22.1	1133	10	O242F3
18	1364	22.1	1124	10	O242F3
19	1357.5	22.0	1103	10	O242F3
20	1353	21.9	1012	10	O242F3
21	1353	21.9	1236	10	O242F3
22	1349.5	21.8	1095	10	O242F3
23	1348	21.8	1008	10	O242F3
24	1346	21.8	1012	10	O242F3
25	1340.5	21.7	1134	10	O242F3
26	1339	21.7	1173	10	O242F3
27	1335	21.6	1064	10	O242F3
28	1332	21.4	1155	10	O242F3
29	1330	21.3	1140	10	O242F3
30	1313	21.2	1013	10	O242F3
31	1310	21.2	1002	10	O242F3
32	1298	21.0	1109	10	O242F3
33	1298	21.0	1110	10	O242F3
34	1293.5	20.9	1092	10	O242F3
35	1291.5	20.9	1098	10	O242F3
36	1281.5	20.7	1178	10	O242F3
37	1279.5	20.7	994	10	O242F3
38	1279	20.7	994	10	O242F3
39	1275.5	20.6	1130	10	O242F3
40	1273	20.6	1123	10	O242F3
41	1262	20.4	1041	10	O242F3
42	1260.5	20.4	1145	10	O242F3
43	1257.5	20.3	996	10	O242F3
44	1256.5	20.3	987	10	O242F3
45	1255	20.3	1002	10	O242F3

## ALIGNMENTS

RESULT 1	ID	Score	Query Match	Length	ID	Description
O22476	1	6140	99.3	1196	10	O22476
O22476	2	3050.5	49.3	1121	10	O242F3
O22476	3	2643.5	42.8	1192	10	O242F3
O22476	4	2633.5	42.6	1164	10	O242F3
O22476	5	2623.5	42.4	1166	10	O242F3
O22476	6	2582.5	41.8	1166	10	O242F3
O22476	7	2509.5	40.6	1143	10	O242F3
O22476	8	2240	36.2	1110	10	O242F3
O22476	9	1608.5	26.0	1110	10	O242F3
O22476	10	1593.5	25.8	1129	10	O242F3
O22476	11	1503	24.3	1120	10	O242F3
O22476	12	1419	23.0	1132	10	O242F3
O22476	13	1402.5	22.7	1106	10	O242F3
O22476	14	1380.5	22.3	1102	10	O242F3
O22476	15	1380	22.3	1079	10	O242F3
O22476	16	1370.5	22.2	1141	10	O242F3
O22476	17	1367	22.1	1133	10	O242F3
O22476	18	1364	22.1	1124	10	O242F3
O22476	19	1357.5	22.0	1103	10	O242F3
O22476	20	1353	21.9	1012	10	O242F3
O22476	21	1353	21.9	1236	10	O242F3
O22476	22	1349.5	21.8	1095	10	O242F3
O22476	23	1348	21.8	1008	10	O242F3
O22476	24	1346	21.8	1012	10	O242F3
O22476	25	1340.5	21.7	1134	10	O242F3
O22476	26	1339	21.7	1173	10	O242F3
O22476	27	1335	21.6	1064	10	O242F3
O22476	28	1332	21.4	1155	10	O242F3
O22476	29	1330	21.3	1140	10	O242F3
O22476	30	1313	21.2	1013	10	O242F3
O22476	31	1310	21.2	1002	10	O242F3
O22476	32	1298	21.0	1109	10	O242F3
O22476	33	1298	21.0	1110	10	O242F3
O22476	34	1293.5	20.9	1092	10	O242F3
O22476	35	1291.5	20.9	1098	10	O242F3
O22476	36	1281.5	20.7	1178	10	O242F3
O22476	37	1279.5	20.7	994	10	O242F3
O22476	38	1279	20.7	994	10	O242F3
O22476	39	1275.5	20.6	1130	10	O242F3
O22476	40	1273	20.6	1123	10	O242F3
O22476	41	1262	20.4	1041	10	O242F3
O22476	42	1260.5	20.4	1145	10	O242F3
O22476	43	1257.5	20.3	996	10	O242F3
O22476	44	1256.5	20.3	987	10	O242F3
O22476	45	1255	20.3	1002	10	O242F3

RESULT 1	ID	Score	Query Match	Length	ID	Description
O22476	1	6140	99.3	1196	10	O22476
O22476	2	3050.5	49.3	1121	10	O242F3
O22476	3	2643.5	42.8	1192	10	O242F3
O22476	4	2633.5	42.6	1164	10	O242F3
O22476	5	2623.5	42.4	1166	10	O242F3
O22476	6	2582.5	41.8	1166	10	O242F3
O22476	7	2509.5	40.6	1143	10	O242F3
O22476	8	2240	36.2	1110	10	O242F3
O22476	9	1608.5	26.0	1110	10	O242F3
O22476	10	1593.5	25.8	1129	10	O242F3
O22476	11	1503	24.3	1120	10	O242F3
O22476	12	1419	23.0	1132	10	O242F3
O22476	13	1402.5	22.7	1106	10	O242F3
O22476	14	1380.5	22.3	1102	10	O242F3
O22476	15	1380	22.3	1079	10	O242F3
O22476	16	1370.5	22.2	1141	10	O242F3
O22476	17	1367	22.1	1133	10	O242F3
O22476	18	1364	22.1	1124	10	O242F3
O22476	19	1357.5	22.0	1103	10	O242F3
O22476	20	1353	21.9	1012	10	O242F3
O22476	21	1353	21.9	1236	10	O242F3
O22476	22	1349.5	21.8	1095	10	O242F3
O22476	23	1348	21.8	1008	10	O242F3
O22476	24	1346	21.8	1012	10	O242F3
O22476	25	1340.5	21.7	1134	10	O242F3
O22476	26	1339	21.7	1173	10	O242F3
O22476	27	1335	21.6	1064	10	O242F3
O22476	28	1332	21.4	1155	10	O242F3
O22476	29	1330	21.3	1140	10	O242F3
O22476	30	1313	21.2	1013	10	O242F3
O22476	31	1310	21.2	1002	10	O242F3
O22476	32	1298	21.0	1109	10	O242F3
O22476	33	1298	21.0	1110	10	O242F3
O22476	34	1293.5	20.9	1092	10	O242F3
O22476	35	1291.5	20.9	1098	10	O242F3
O22476	36	1281.5	20.7	1178	10	O242F3
O22476	37	1279.5	20.7	994	10	O242F3
O22476	38	1279	20.7	994	10	O242F3
O22476	39	1275.5	20.6	1130	10	O242F3
O22476	40	1273	20.6	1123	10	O242F3
O22476	41	1262	20.4	1041	10	O242F3
O22476	42	1260.5	20.4	1145	10	O242F3
O22476	43	1257.5	20.3	996	10	O242F3
O22476	44	1256.5	20.3	987	10	O242F3
O22476	45	1255	20.3	1002	10	O242F3

Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL; AF017056; AAC49810.1; -.  
 DR EMBL; AL078620; CAB4675.1; -.  
 DR EMBL; AL161595; CAB80603.1; -.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR003592; LRR\_out.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR InterPro; IPR004040; STY\_pkinase.  
 DR Pfam; PF00560; LRR; 20.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR00019; LEURICHRPT.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00370; LRR; 15.  
 DR SMART; SM00221; STYK; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR ATP-binding; Serine/threonine-protein kinase; Transferase.  
 SEQUENCE 1196 AA; 130542 MW; C7FBA1C21294E600 CRC64;

Query Match 99.3%; Score 6140; DB 10; Length 1196;  
 Best Local Similarity 99.4%; Pred. No. 0;  
 Matches 1189; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 MKTSSFFLSTVTLFFSFPSLSFQASPSQSLYREIHQILSFQDVLDPKXLLPDMSSNNX 60  
 Db 1 MKTSSFFLSTVTLFFSFPSLSFQASPSQSLYREIHQILSFQDVLDPKXLLPDMSSNNX 60

QY 61 PCTFDGVTCTRDVKYTSIDLSKRPINVGFSVSSLSLTGLESPLFNSHINHSVSQFRC 120  
 Db 61 PCTFDGVTCTRDVKYTSIDLSKRPINVGFSVSSLSLTGLESPLFNSHINHSVSQFRC 120

QY 121 SASLTSIDLSNSISGPTVTLTSLGSCGKLFNVSSNTLDPKGVSGGLKXLSLEVL 180  
 Db 121 SASLTSIDLSNSISGPTVTLTSLGSCGKLFNVSSNTLDPKGVSGGLKXLSLEVL 180

QY 181 SANSISGANVGVWVLSDCGELKHLAISGNKISGDVIVRCVNIPLDVSSNNSTGIFP 240  
 Db 181 SANSISGANVGVWVLSDCGELKHLAISGNKISGDVIVRCVNIPLDVSSNNSTGIFP 240

QY 241 LGDGCALDHLDSGNKLSGDSFRAISTCTELKLNISNOFVGPPLPLSLQYLSLAE 300  
 Db 241 LGDGCALDHLDSGNKLSGDSFRAISTCTELKLNISNOFVGPPLPLSLQYLSLAE 300

QY 301 NKFTGEIPDLFSGACDTLTGLDLSGNHFYGANPPFGSCSLLESIALSSNNFSGELPMDT 360  
 Db 301 NKFTGEIPDLFSGACDTLTGLDLSGNHFYGANPPFGSCSLLESIALSSNNFSGELPMDT 360

QY 361 LKMGGLKVLDSFNEPSGELPESLITNLSASLLTDLSSNNSGTLINLQNPNTLQ 420  
 Db 361 LKMGGLKVLDSFNEPSGELPESLITNLSASLLTDLSSNNSGTLINLQNPNTLQ 420

QY 421 LYLQNNGFPGKLPPLTNSGSELVSLHSFNYLSGTLPSLSGLSKLRDLKLTMLNLEGEI 480  
 Db 421 LYLQNNGFPGKLPPLTNSGSELVSLHSFNYLSGTLPSLSGLSKLRDLKLTMLNLEGEI 480

QY 481 POELMVVKTLETLIDPNDLTGEIPSGLSNCTNLNWLISLNNRLTGEIPLKWLIENTLAI 540  
 Db 481 POELMVVKTLETLIDPNDLTGEIPSGLSNCTNLNWLISLNNRLTGEIPLKWLIENTLAI 540

QY 541 LKLSNNSFSGNIPDELGDGSLIWLIDNLNLRGTIPAMFQSGKLIANFAGRYVYI 600  
 Db 541 LKLSNNSFSGNIPDELGDGSLIWLIDNLNLRGTIPAMFQSGKLIANFAGRYVYI 600

QY 601 KNDGKKECHGAGNLLEFQGISREQLNRLSTRPNCITSRVYGHTSPFDNNGSNMPLD 660  
 Db 601 KNDGKKECHGAGNLLEFQGISREQLNRLSTRPNCITSRVYGHTSPFDNNGSNMPLD 660

QY 661 MSYNMLSGYIPKEIGSNPYLFIINLGHNDISGSI PDEVGLRGLNLTLDLSNKLGRIRIQ 720  
 Db 661 MSYNMLSGYIPKEIGSNPYLFIINLGHNDISGSI PDEVGLRGLNLTLDLSNKLGRIRIQ 720

QY 721 AMSALTLTLEIDLSNNNLGSPILPEMGOFETPPAKFLNNGPGLCYPLPCDCPSNADGYAH 780  
 Db 721 AMSALTLTLEIDLSNNNLGSPILPEMGOFETPPAKFLNNGPGLCYPLPCDCPSNADGYAH 780

QY 781 HQRSHGRPASIAGSVAMGLIFSFVCI FGLILVGRBMRKRRRKEALEMYAEGHNSGD 840  
 Db 781 HQRSHGRPASIAGSVAMGLIFSFVCI FGLILVGRBMRKRRRKEALEMYAEGHNSGD 840

QY 841 RTANNTWKLTGYKEALISITLAAPEKPLRLTADLLQATNGHNLSLIGSGGFQGVYKA 900  
 Db 841 RTANNTWKLTGYKEALISITLAAPEKPLRLTADLLQATNGHNLSLIGSGGFQGVYKA 900

QY 901 LKDGSAVALIKLIVHSGQGDREPMEMETIGKI KHRNLVPLIGYCKVGBERLLVYEFMK 960  
 Db 901 LKDGSAVALIKLIVHSGQGDREPMEMETIGKI KHRNLVPLIGYCKVGBERLLVYEFMK 960

QY 961 YGSLLEDVLPDPKKGVLKLTSTRKKAIGSARGLAFIHNCSPHIIRDMKSNVLLDEN 1020  
 Db 961 YGSLLEDVLPDPKKGVLKLTSTRKKAIGSARGLAFIHNCSPHIIRDMKSNVLLDEN 1020

QY 1021 LEARVSDPGMARLMSADTHLSVSTLACTGVPPEYVYOSFRCSTKGDVYSYGVLLLEL 1080  
 Db 1021 LEARVSDPGMARLMSADTHLSVSTLACTGVPPEYVYOSFRCSTKGDVYSYGVLLLEL 1080

QY 1081 TGKRPDSDPDGDNLLVGVWQKAKLRISDFPELMEKEDPALIEILLQILKVAACLDD 1140  
 Db 1081 TGKRPDSDPDGDNLLVGVWQKAKLRISDFPELMEKEDPALIEILLQILKVAACLDD 1140

QY 1141 RAMRRPTVQVAMFKEIQAGSGIDSGSTIRSIDGSPSTLEMDMSIKVPEEKL 1196  
 Db 1141 RAMRRPTVQVAMFKEIQAGSGIDSGSTIRSIDGSPSTLEMDMSIKVPEEKL 1196

RESULT 2  
 ID 0942F3 PRELIMINARY; PRT; 1121 AA.  
 AC 0942F3;  
 DT 01-DEC-2001 (Tremblrel, 19, Created)  
 DT 01-DEC-2001 (Tremblrel, 19, Last sequence update)  
 DT 01-MAR-2002 (Tremblrel, 20, Last annotation update)  
 DE Putative brassinosteroid-insensitive protein BR1.  
 GN P0480C01.13.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;  
 OC Eriocarpaceae; Oryzaceae; Oryza.  
 OC NCBI\_TaxID=4530;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
 clone: P0480C01.1";  
 RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AP003453; BAB68053.1; -.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR Pfam; PF00560; LRR; 19.  
 DR Pfam; PF00069; pkinase; 1.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; UNKNOWN\_1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; UNKNOWN\_1.  
 KM ATP-binding; Transferase.  
 SQ SEQUENCE 1121 AA; 120180 MW; F71A49845B0E2D09 CRC64;

Query Match 49.3%; Score 3050.5; DB 10; Length 1121;  
 Best Local Similarity 53.9%; Pred. No. 1.2e-190;  
 Matches 632; Conservative 166; Mismatches 276; Indels 99; Gaps 20;

QY 39 LISFQDVLDPKXLLPDMSSNNKCTFDGVTCTRDVKYTSIDLSKRPINVGFSVSSLSLT 98

```

Db 31 LEEFRQAVNPQAAKKGSGDGA CRFPAGCRNGRLTSLSLAGVPLNABRAVAATLQL 90
Qy 99 TGELEFLSNHNGSVS---GFKCSALSTLDSRN-SLSGPTTLTISG- CSQLKFL 153
Db 91 GSVEVLRLRANVSGALSAAGARCSGLQHLDSGNALRGSVADVAALASACGGKCTL 150
Qy 154 NVSSNTLDLPFGKV--SGGLKLNLLEVLDSANSISGANVVGWYLDGCGELKHLAI SGNK 211
Db 151 NLSGDAVG-AAKVGGGGGPGAGLDSLDSNNKLTDSDLRMWVADGAVRMLDLALNR 209
Qy 212 ISGVDVSRVCMLEFLDVSSNNSTGTLPFGDCSALQHLDISGNKLSGDF-SRAISTCTE 270
Db 210 IS-----GVPEFTNCSGLQYLDLSGNLIVGEVPGALSDCG 246
Qy 271 LKLNLSNQFVGPIPLPLKSLQYLSLANKEGTGELPDLISACTLTGLDLSGNHFG 330
Db 247 LKLN-----LSFNHLG 259
Qy 331 AVPPFGSCSLLESLSLSSNNFSGELPMDTLKMRGLKVLDSLFNEFSGELPSTLNTLSA 390
Db 260 VFPEDLAGLSTLANLSSNNFSGELPGEAFKQLQGLTALSFPNPNSSIPPTVASL-P 318
Qy 391 SLTLTDLSSNNEGPIPLNLCNPKNTLOELYQNNGFTGKLPPTLSNCSSEVSLHSTFN 450
Db 319 ELQOOLDSNTFSGITPSSLCODPNKTLHLTYQNNVYLTGPIPAVSNCTSLVSLDLSLN 378
Qy 451 YLSGTPSSLSGLSKLRLKLMLEGEIPELMVYKTLFTLIDPNULTIGEISGLSN 510
Db 379 YINGSLPASLGDIGNLODLTLWONELEGETPAISRIQGLEHLIDYNGLTGSIPELAK 438
Qy 511 CTNNANISLNNELTGEIPEKMLENLALIKLSNNSFSGNIPDELGDGRLSLIMDLNTN 570
Db 439 CTNNANISLNNELTGEIPEKMLENLALIKLSNNSFSGNIPDELGDGRLSLIMDLNTN 498
Qy 571 LFNGTIPAMFKSGKIAPNFAAGKRYVYIKDDMKKECHGAGNLEFQIRSEQLNRLS 630
Db 499 QLNGLSIPKEILAKSGKQNVGLIVGRPYVYLRNDELSECKRGKSLLEFISIRPDLSSRP 558
Qy 631 TRNPKNTSRVYGHTSPFEDNNGSMFEDMSTNMLSGYIPKISGMPYFLTLNGHNDI 690
Db 559 SKKLCPFT-RMYVGSSTEVFNKKSIMIFLDLSYQNDASIPBELGMFYLMNIGHNLL 617
Qy 691 SSGIPDEVLDLGNLTLDSNNKLDGRIPQMSALTMLREIDLSNNNLSGPIPEMGQFET 750
Db 618 SGTIPRLAEAKKLAVLDLSYQNBSPINPSALS-LSEIMLSNNQNLGTTPELSLAT 676
Qy 751 FPPAKFLNPGLCGYLPRCDPSNADGVANHQSHGRPRPASLAGSVAMGLFSPFCIFGL 810
Db 677 FPKSGYENNTGLCGFPLPCDHSPPRSSNDHQ-SH-RQASMASSILMGLFSLFCIIVT 734
Qy 811 ILVGRBMRGRKKKEAEI-EMVABSHGNSGDRYANTNTWK-LTGVEALSSINLAEEK 866
Db 735 IIAISGKRRLKNBEASTLRDIYIDRSHS--ATMNSDMRNLGSGT-NLISINLAEEK 790
Qy 867 PLRKLTFADLQATNGFNHDSLIGSGFGDYKATILKQSSAVALIKLHVSQSGRETEMA 926
Db 791 PLQNTLADLVATNGFHALCOISGGGFEDVYKAQKQKVAIKKLHVSQSGRETEMA 850
Qy 927 EMEITIGIKHRLVPLLGCKVDERLVNEVMKYGSLEVDLPKKGAVKLKLSSTREKI 986
Db 851 EMEITIGIKHRLVPLLGCKVDERLVNEVMKYGSLEVDLPKKGAVKLKLSSTREKI 910
Qy 987 AIGSARGLAFLHNSGPHIIRHDMKSNVLDLNTENAVVSDFGMARLMSANDTLVSTL 1046
Db 911 AAGARGLAFLHNSGPHIIRHDMKSNVLDLNTENAVVSDFGMARLMSANDTLVSTL 970
Qy 1047 AGTPGVPEPEYQSRGSGKGVYGVVTLLELTGKRPDSPDG-DNNTVGVTKQAK 1105
Db 971 AGTPGVPEPEYQSRGSGKGVYGVVTLLELTGKRPDSPDG-DNNTVGVTKQAK 1030
Qy 1106 LKISVDFELMKEDPALEIELQLKVAVALCLDDRARRRPTMVMFKEIOAGSGSID 1165

```

```

Db 1031 LKTDVPELLEKEDPSVELELLEHLKIAACLDPRSRRPTMKVAMFKEIOAGSTVD 1090
Qy 1166 SOST--IRSIEDGGSFTTEMVDSIKVEPGK 1195
Db 1091 SKTSSAAGSIDEGGYG---VLDMPLEAKEEK 1120

RESULT 3
Q9ARC8
ID Q9ARC8 PRELIMINARY; PRT; 1192 AA.
AC Q9ARC8;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical 129.9 kDa protein.
OS Lycopersicon esculentum (tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RX SEQUENCE FROM N.A.
RP MEDLINE=21178822; PubMed=11283350;
RA Rosenberg M., Theres K., Acarkan A., Herrero R., Schmitt T.,
RA Schumacher K., Schmitt G., Schmidt R.,
RT "Comparative sequence analysis reveals extensive microcolinearity in
RT the lateral suppressor regions of the tomato, Arabidopsis and Capsella
RT genomes."
RL Plant Cell 13:979-988(2001).
CC -I- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL, AJ303345; CAC36401.1; -.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR002290; Ser thr pkinase.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00560; LRR; 21.
DR PRINTS: PR00019; LEURICHRPT.
DR PRINTS: PR00109; TYRKINASE.
DR PRODOM: PD000001; Euk_pkinase; 1.
DR SMART: SM00220; S_TKc; 1.
DR SMART: SM00219; Tyrc; 1.
DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE: PS00111; PROTEIN KINASE DOM; 1.
DR PROSITE: PS00108; PROTEIN KINASE ST; 1.
KW ATP-binding; Hypothetical protein; Serine/threonine-protein kinase;
KW Transferase.
SQ
SEQUENCE 1192 AA, 129941 MW, BDELCEDAF8930886 CRC64;

Query Match 42.8%; Score 2643.5; DB 10; Length 1192;
Best local similarity 49.5%; Pred. No. 5.2e-164;
Matches 573; Conservative 193; Mismatches 341; Indels 51; Gaps 27;

Qy 35 EIHQLSFK-DVLPDKN-LLPWS-SNKNPCTFEDVTRDKVTISIDISSKPLNVGSA 90
Db 45 QVGRLLAFKKSVEVDPNGLNEMTILSSSPCTWNGISCSNGVLAINDS---VGLSG 100
Qy 91 V--SSLSLTGLESFLSNHNGSVSGFKCASLTSLDSRNLSGVTTLTISGSCS 148
Db 101 LHLTDLMLPTLLRNVFSGNHFYGNLSSIASCSPEFLDLSANFSEVLVLEPLKSCD 160
Qy 149 GLKFLNVSSNTLDPPKVSQGLKLN-SLEVLDSANSISGANVVGWYLDGCGELKHLAI 207
Db 161 NIKLVNLSGNSI--KGVV--LKFGPSLQDLDSNTISIDPGLSYALSN-CQNTANLNF 214
Qy 208 SGNKISGV--DVRVCMLEFLDVSSNNFSTGIP-FLGDCSALQHLDISGNKL-SGDFS 262
Db 215 SSNKGKLSKSSISCKSSVLDLSRNNTLGEINLIDLGTCQNTLVNLSFNNTLSVERP 274
Qy 263 RAISTCEIKLNISSNQVGPFP---LPLSKQYLSLAENKFTGELPDLISGACTLT 319
Db 275 PSLANCSLNTANINANSIRMEIPVELLVKLSKRLVLAHNOFPDKIPSELGSCSTLE 334

```

Qy	320	GLDLSGNHFGAVMPFRFGSCSLLESLATSSNNFSEGLMDTLMMRGKYLDSFNEFS	379
Db	335	ELDLSGNRLTGLPESTFKCSSLPSTLNGNNLSDDPLATWISSITLRLTYLPENNITG	394
Qy	380	ELPESTLNLASLTLTDLSSNNFSGPILNLCQNPKN-TLQELYLQNNFTGKIPTPLSN	438
Db	395	YVPSKLVNCT-KLQVLDLSSNAFIGNVPSEFCFAAGPFLFMTLLASVYLGTVPKOLGH	453
Qy	439	CSELVSLHLSFNVLSTGISPSLGSLSKLPDLKLTMLMEGLPEOL-MYVTTLETLLDF	497
Db	454	CRLNRKIKDLSFNNLVQSIPIELIWNLPNLSELMWANNLTGELPBCICINGNLQTLILAN	513
Qy	498	NDLTGEIPSGISNCTNLMTWISLNNNLGTSEIPKWIQLENTALIKLSNNSPSGNI	557
Db	514	NFISGTLPOGISKCTNLVWVWISLSSNNLSSEITQGGIGNLNLATLIQDGNNSLTGPI	573
Qy	558	DCRSLIMDLNLTNLTENGTLPAAMFKOSGKIANTFLAGRYVYITKIDGKKKECHGAGNLE	617
Db	574	SCRNLIMDLNLSNLTNLTESIPLLELADQAGHVNPMASGKQFAFVNRBG-GTCRAGAGLVE	632
Qy	618	FOGIRSGQLNRSLNRNCTNLTSSVVGQHSPPFDNNGSWMFQDMSSYNNLSGYIPKEISGM	677
Db	633	FEGRREELALIPNVHCCBPT-RIYSRRTYTTSSGSIYIYDLSYNLSSEITPDNLSGL	692
Qy	678	PYLFILNLGNDISGSIPEVEVDLRGLNLIIDSSNKLDRIPQAMSALTMLTEIDLSNN	737
Db	692	SFLQVNLNGHNNFPTGPIPFNFGGIKITGVGLDLSHNSLQGFIPPSLGSLSFLSDLDVSN	751
Qy	738	LSGPIPEMGQETPEPPAKFLFNNPGLCGVPLPRCDPSNADGVAMHQRSGQRAPSLAG	794
Db	752	LSGTLPSGGQLTTPPARVYENNSLCCVPLPCCSGSNGHSSSIYHGNKKPTTIG	807
Qy	795	SVAMGLTSPFCIFGLILVGREMKRRKKKAELEMYABGHGNSGDRANTNMKLVGK	854
Db	808	MV-VGIMVSPICILLVIALYKIKKQNEBE-KRDXYIDSLPTSGSSSMWLSVTP	860
Qy	855	EALSTNLAAFEKPLRKLTLPADLLQATNGHNDLSLSSGSGDVYKALIKDGAVALIKLI	914
Db	861	BPLSLNVAITPFKPLRKLPFGHLEATNGFSSSEMSIGSGGEYTKQKLDGSLVIAIKLV	920
Qy	915	HVSGGQDEPMAEMETTGKIGHRLVPLGYCKVQGBERLLVNEVMKYGSLABVLQDPKKG	974
Db	921	HVTGGQDEPMAEMETTGKIGHRLVPLGYCKVQGBERLLVNEVMKYGSLABVLQDPKKG	980
Qy	975	GVKLKLTSTRKTIAGSARGALFLLHNSCPHIIHDKMSNVLLDENLEKRVSDFGMALM	1034
Db	981	GMFLDMPARKKIAIGSARGALFLLHNSCPHIIHDKMSNVLLDENLEKRVSDFGMARLV	1040
Qy	1035	SAMDHLTASVTLAGPGVVPPEYQSFSCSKGQVSYGVVLLLELTGTGRPTDSDPFG-D	1093
Db	1041	NMLDHLTASVTLAGPGVVPPEYQSFSCSKGQVSYGVVLLLELTGTGRPTDSDPFGD	1100
Qy	1094	NNLVGVWVKQ-HAKLAIIDVFPPELMEKEDPALEIFELLQHLKVAACIDDRAMRRPTWQVM	1155
Db	1101	NNLVGVWAKQJLHNDKSHIELDPELT-TVLSGDALVTHYLLKVAECIDERSYKRPWTQVM	1155
Qy	1153	AMFKEIQAGSGIDSOSTI 1170	
Db	1160	TKFEVQT---DSESDI 1173	
RESULT 4			
Q9LJF3 PRELIMINARY; PRT; 1164 AA.			
AC	Q9LJF3;	01-OCT-2000 (TREMBlrel. 15, Created)	
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)		
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)		
DE	Receptor protein kinase.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi		

[illegible]



```

Db 474 GANETTLNNNNLITGSI:PESTISCKTNMLWTSLSNLLTGELIPGICKLELTLQIGNN 533
QY 547 SPSGNIDPELDCRSLIWLIDNTNLFNGLTIPAMFKOSGKIAPAFIAGKRYVYKJNGM 606
Db 534 SLTGNIPSELGNCNLLIWLIDNSNLLTGNLPGELASQALVPSVSGKOPAFKPNBG-G 552
QY 607 KECHGAGNLEFQGIQISEQLNRLSTRNPNCTITSRYVGHSTSPFDNNGSNMFLDMSYNL 666
Db 593 TDCGAGGLVEFEGIRARERLEHFPVHVSCEPT-RISGMMYMPSSNGSMIYLDLSINAV 651
QY 667 SGYPKSIGSMFYFIINTLGHNDISGIPDEVDLRLNLIDLSNLTDRIGQAMSLT 726
Db 652 SGTSLPLGAGMGYQVANTLGNLITGTLPSFGGLKAIQVLDISHNDLQGLPGLSGLS 711
QY 727 MLTEIDLSNNNLSGPIDEMQOFETFPAPKFLNNGLCYGLPCDPSNADYAHQSHG 786
Db 712 FLSDLDVSNNNLTGPIDFGGLTFPLTRYANNGLCGVLLPCSSG-----SRPRSHA 766
QY 787 R-RPASLAGSAMGLLSPFCIFGLIYGRBMRKRKKEAELEMTAEHSGNSGDRTPANN 845
Db 767 HPKQSTATATGASGIVSFMCITWLIWALYRARK-VQKKEKREKYLESLPTSG-----S 820
QY 846 TNNKLTGVEKALSTINLAFAEKPLRKLTFADLQATNGFNHDSLISGGGEGDYKALKKG 905
Db 821 SSMKLSVHEPLSINAVATFEKPLRKLTFADLQATNGFNHDSLISGGGEGDYKALKKG 880
QY 906 SAVAIAKLLHVSQGGDEFEPAEMETIGIKIRNMLVPLLYGCKYDERRLLVNEVMKYSLE 965
Db 881 SVAIAKLLIQTGGDEFEPAEMETIGIKIRNMLVPLLYGCKYDERRLLVNEVMKYSLE 940
QY 966 DVLQD-PRKSGVTLKASTRKIAIGSRGLAFIHNHCSPIIHRDKSSNVLIDENLEAR 1024
Db 941 TVLHEKTKGKGTFLDMSARKKIAIGARGLAFIHNHCSPIIHRDKSSNVLIDQDFVAR 1000
QY 1025 VSDPGRALMSANDTHLSVTLTLAGTGYVPEYXQSFRCSTKGDVSYGVILLELTGR 1084
Db 1001 VSDPGRALMSANDTHLSVTLTLAGTGYVPEYXQSFRCSTKGDVSYGVILLELTGR 1060
QY 1085 PMSDPFG-DNNIVGWVKQ-HAKLRISDVDFPELMKEDPALTEILLQKVAACLDLRA 1142
Db 1061 PLDPSEFGEEDNNLVGWAKOLYREKRAEILLPELV-TDKSGVIELLHYLKINSQCLDHP 1119
QY 1143 WRSPYVOWAMAKE-IOAGSGIDS 1166
Db 1120 FKRPYMIQVMMFKELVQVDTENDS 1144
QY 5
Db 5
AC 09ZWC8 PRELIMINARY; PRT; 1166 AA.
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE F2002.4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosida 11; Brassicales; Brassicaceae; Arabidopsis.
OK NCBI_TaxID=3702;
RN
RP SEQUENCE FROM N.A.
RA Khan S., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Kim C.,
RA Shinn P., Altafi H., Bel Q., Chin C., Chou J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharbeky N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RA "Genomic sequence for Arabidopsis thaliana BAC F20N2 from chromosome
RT 1."
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

```

```

RN [2]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Shinn P., Buehler E., Dunn P., Feng J.J., Kim C.C., Li Y.Y.,
RA Walker M.W., Altafi H.H., Araujo R.R., Conn L.L., Conway A.A.B.,
RA Gonzalez A.A., Hansen N.N.F., Huizart L.L., Kremenetskaia I.I.,
RA Lenz C.C., Li J.J., Liu S.S., Luros S.S., Rowley D.D., Schwartz J.J.,
RA Toriumi M.M., Vysotskaya V.V., Yu G.G., Davis R.R.W.,
RA Federspiel N.N.A., Theologis A.A., Ecker J.R.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bel Q., Chin C., Chou J., Choi E., Conn L.,
RA Conn L., Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharbeky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
CC -1 SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AC002328; AAF79510.1; -.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003592; LRR_out.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR004040; STY_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00560; LRR; 21.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00019; LEURICHRPT.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00370; LRR; 16.
DR SMART; SM00221; STYKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Serine/threonine-protein kinase; Transferase.
SQ
SEQUENCE 1166 AA; 127423 MW; 8C4DD9231A466A7 CRC64;
Query Match 42.4%; Score 2623.5; DB 10; Length 1166;
Best Local Similarity 48.5%; Pred. No. 1e-162;
Matches 573; Conservative 190; Mismatches 341; Indels 77; Gaps 29;
QY 33 YREIHQISFK-DVLPD-KNLLPDW--SSNKNPCTFDGVTGCD-KVTSIDLSKPLNV 86
Db 32 FNEFALLAKKNSVSDPNNTVGNMKYSGKSCSWRGVSCDDRIYGLDIRNSGLTG 91
QY 87 GFSAVSSSLITGLSPLNSHIN-GVSGFKCSASITSLDLSRNLGSPVTLTSL 144
Db 92 TLNIV-NLTALPNLQNLVYQGNVFSGGDSGSDC-VLQVLDLSNSISIDYSNWDYF 147
QY 145 GSGSGLKFNVSNTLDFGKYV-GGLKINSLEVDLSANSISGANVGVWVLSGCGEIK 203
Db 148 SKSNLVSVNISNKL-VKQGFAPDSQSILTYVDSLSTNLSL-KIPRSFISDPASLK 204
QY 204 HLAISNKISGVDVRCVNLLEFLDVSNNSFTGIPFLDGCALQHLDSGKLSGD-PS 262
Db 205 YLDLTHNNLSG-----FSDLS-----FQICGNLTFPSLSQNNLSGDKFP 244
QY 263 RAISTCTELKLNISNQVGFIPP---LPLKSLQVLSLAENKFTGETIPDLGACDTL 318

```

Db 245 ITLPNCKFLFTLNISRNKLAKIPNGRWYGSFONLKQSLAHNRISGEIPPELSLCKTL 304  
 Qy 319 TGLDLSNNHFGAVPPFGSCSLLESALSSNNFSGELPMDTLKRGKLVLDLSNERS 378  
 Db 305 VIIDLSSNTGESGELPSQFTACVWLQNLNLSNLSGFLNTVSKTGTITLYLVATNNTS 364  
 Qy 379 GELPESTLNLSASLLTLDLSNNSGFLTLPNLCQ-NPKNTLQELIYLNONGFTKIPPTLS 437  
 Db 365 GSVISLTLNCS-NLRVLDLSNGFTGVPSGFCSSQSPVLEKLIANNVLSGVPELQ 423  
 Qy 438 NCSEVLSLHSENYLSGTPSSLSLSKLDKIMLMLSEIPEQELMYK--TLETLLI 435  
 Db 424 KCKSLKTLIDLSFNLGTGIPKEIMWLPLSLVMMANNLTGTLIP-EGVCYKGGNLTETLL 482  
 Qy 496 DPNLDTGEIPEGSLNCTNMLTISLNNRGTGEIPKMGLEMLAILLKSNNFSGNIPDE 555  
 Db 483 NNNLITGSIPISSIRCTNMWISLSSNLTGKIPSGIGLSKILLQGNNSLSGNVPRQ 542  
 Qy 556 LGDCRSLLTMDLNTNLTNGTIPAMAFKQSGKIAANFLAGKRYVYIKNDGMKKECHGAGNL 615  
 Db 543 LQNCCKSLTMDLNSNNLTGDLPEGLASQAGLVMPGSYGQFAFVRNMG-GTDCRAGGL 601  
 Qy 616 LEFGQIRSEQLNRSLTRNCPNITSRVYGGHTSPPTDNNNGSMFLDMSNMLSGTYIPKEIG 675  
 Db 602 VEFEGIRAEERLERLPMVHSCPAT-RYSGMTMYTFPSANGSMIYEDISYNAVSGEIPEGYG 660  
 Qy 676 SMPYLFITNLGNNDISGSIPEVGDRLNLTLDLSNLTDLRIQAMSLTMLTEIDLSN 735  
 Db 661 NNGYLVQVNLGNRLTGTIPDSFGELKAIQVLDLSHNLQGLVGSLSLSFLDLVDVN 720  
 Qy 736 NNLSPGIDEMGQFETFPAPKFLANNPGLCGYPLPRCDPSNADGVAHQRSGRRPASLAGS 795  
 Db 721 NNLTPPIRPGQGLTFPYSRYANNGLCGVPLRPC--GSAPRRPITSRIHAKK-QYVATA 777  
 Qy 796 VAMGLIFSVCLFGLILVGRERKRRKKEALEMYAGHSGNSGDRANTNMNLTGKVE 855  
 Db 778 VTAGIAFSFMCVVMVLMYVRK-VQKEQKREKITSLSLPSG-----SCSWKLSSVPE 831  
 Qy 856 ALSINLAERKPLRKLTFADLLQATNGFNDLSLISGGFGDYKAILKDGSAVAIKKLIH 915  
 Db 832 PLISINVAITEKPLRKLTFADLLQATNGFNDLSLISGGFGDYKAILKDGSAVAIKKLIH 891  
 Qy 916 VSGQGDREFMAEMETIGIKIHNLYVLLGYCYGVDRLIVNEVMKYSLEDVLDQ--PKK 973  
 Db 892 ITGGQGRREFMAEMETIGIKIHNLYVLLGYCYGVDRLIVNEVMKYSLEDVLDQ--PKK 951  
 Qy 974 GGYKLTSLTRKRIKAIQSAGLAFLHNCSPHIIRPMKSSNVLLDENLEARYSDGMARL 1033  
 Db 952 GGYIYLMARKKAIQAAGLAFLHNCSPHIIRPMKSSNVLLDENLEARYSDGMARL 1011  
 Qy 1034 MSAMDTLSTVSTAGTPGYVPEPYQSFRCSTKGDYSGVVLLELTGKRPDPSDFG- 1092  
 Db 1012 VSLDLHLSTVSTAGTPGYVPEPYQSFRCSTKGDYSGVVLLELTGKRPDPSDFG- 1071  
 Qy 1093 DNNLVGVWVQO-HAKLRISDVPELMEKEDPALELLOHLKVAVACLDLRARRPRMVOY 1151  
 Db 1072 DNNLVGVWVQO-HAKLRISDVPELMEKEDPALELLOHLKVAVACLDLRARRPRMVOY 1130  
 Qy 1152 MAMPEKIQAGSGIDSGTIRSIEDGGSSTEMVDMISKEVP 1192  
 Db 1131 MAMPEKIQAGSGIDSGTIRSIEDGGSSTEMVDMISKEVP 1156  
 RESULT 6  
 O9ARF3 PRELIMINARY; PRT: 1166 AA.  
 AC O9ARF3;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)  
 DE Hypothetical 127.3 kDa protein.  
 OS Capsella rubella.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Capsella.  
 OK NCBI\_TaxID=81985;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21178822; PubMed=11283350;  
 RA Rosenberg M., Theres K., Aarhan A., Herrero R., Schmitt T.,  
 RA Schumacher K., Schmidt G., Schmidt R., extensive microcolony in  
 RT "Comparative sequence analysis reveals the lateral suppressor regions of the tomato, Arabidopsis and Capsella  
 RT genomes.";  
 RL Plant Cell 13:979-988 (2001).  
 CC 1. SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL, AJ030349; CAC6390.1; --  
 DR InterPro; IPR000719; Euk\_kinase.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR002290; Ser\_thr\_kinase.  
 DR InterPro; IPR001245; Tyr\_kinase.  
 DR Pfam; PF00560; LRR; 18.  
 DR Pfam; PF00063; kinase.1.  
 DR PRINTS; PR00019; LEURICHRP.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_kinase; 1.  
 DR SMART; SM00220; S\_trc; 1.  
 DR SMART; SM00219; TyrKc; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR ATP-binding; Hypothetical protein; Serine/threonine-protein kinase;  
 KW transferase.  
 SQ SEQUENCE 1166 AA; 127261 MW; DADE2E2B4C5647A0 CRC64;  
 Query Match 41.8%; Score 2582.5; DB 10; Length 1166;  
 Best Local Similarity 47.2%; Pred. No. 4.9e-160; Indels 103; Gaps 30;  
 Matches 576; Conservative 188; Mismatches 353;  
 Db 8 FLGVTLTFFSFFSLSPQASPSQSYREIHQISFK-DVLPD-KNLLPDW--SSNKNPC 62  
 5 WLFVVLICFFTLGHHGKRL-INSDFEFLALMAFQFQSVKSDPNVVGWVYEGSGRGSC 63  
 Qy 63 TPDGVTCRDD-KVTSIDLSKPLNFGSAV--SSLSLSTGLESPLNSHINGSVSGK 119  
 Db 64 SWRGVSCSDGDRIGVLDLR---NGGVGTGLNMLANTALPNLQNYLYQNVSSSSSGGS 119  
 Qy 120 CSAS--LTSIDLSRSLSGPTVTLTSLSCSGELKPLNVSNTLPPGKVS--GGLKLNLSL 175  
 Db 120 SSGSYCYLQVLDLSNLSIDYSLVDYVFSKSNVSVNFSNNKL--VGKLGAPSSLSKL 177  
 Qy 176 EVLDSANSISGANVVGWVLDSCGELKHLAISGNKISG--DVDVSRCVNLEFLDVSN 232  
 Db 178 TYVDFSNILS-EKIPESFISEFPASLKTYLDLTHNNSGDFSLSPGMCNLSFFSLSON 236  
 Qy 233 NFSTGIPFLDCALQHLDISGNKLSGDPRAISCTELKLNISNQVGPVPP---L 288  
 Db 237 N-----ISGVK---FPISLNCRFETLNLNNAKIPGGEYWG 274  
 Qy 289 PLKSIQVLSAENKFTGEIPDFLSGACDTLTGDLISGNHFGAVPPEFSCSLLESALS 348  
 Db 275 SFQNLKQSLAHNFSEIPELSLCKITLETLDLSGNALSGLPQFACVWLQVNLNG 334  
 Qy 349 SNMFSGELPMDTLKMGKLVLDLSFNEFSGELPESITNLASLLTLDLSNNFSGPILP 408  
 Db 335 NNYLSGDFLSTVSVSKIRITLYLVAFNNISGVISITNCT-NLRVLDLSNSNGFTGVPS 393  
 Qy 409 NLQ-QNPKNTLQELIYLNONGFTGKIPTTNSCEIYLSHSFNYLSGTPSSLSKLR 467  
 Db 394 GLCSQSSPVLEKLIANNYLSGTVPELCKSKSLKTLTETDLSGNALSGLPQFACVWLQVNLNG 453  
 Qy 468 DLKLMNMLEGEIPEQELMYK--TLETLLIDPNLDTGEIPSGSNCTNMLWISLNNRLT 525  
 Db 454 DLVMMANNLTGSIPE-EGVCYKGGTLETTIILNNTLITGSIQGISRCNMIMWISLNNRLT 512

QY 526 GEIPKTIARLENTAIIKLSNNSSGNI PDELGDRCSLIIMDLNTNLNFTIPAMKQSG 585  
 Db 513 GKITPGISNLSKIALIQLGNNSSLGSNVPROLGCKSLIIMDLNNTLIGDLPGLASQAG 572  
 QY 586 KIAANFIAGKRYVYIKNDGMKKECHGAGNLEFQIRSEDLNLSRNCNTSRVYGGH 645  
 Db 573 LVMGSGVSGQFAVRNEG-CTDRGAGGLVEFEGIAERLEFEPVHSCPAT-RVYSGM 630  
 QY 646 TSPFPDNNNGSMFLDMSYNNMLSGYIPKEIGSMPLFLNIGHNDISGSI PDEVDGDLRGLN 705  
 Db 631 TMYTFSANSGSMYIFDISYNAVSGILPGYGMGYLQVNLGNHRTIGNIPDSIGGLKAI 690  
 QY 706 ILDISSNKLGRIRIQAMASALTMLTEIDLNNNTSGPIPEMGCEPIFPAPKFLNPGLCGY 765  
 Db 691 VLDLSHNDLQGYLPFGSLGSLFSDLDVSNNTLGPPIFGQGLTFPPVRYANNGLCGV 750  
 QY 766 PLPRCDPSNADGYAHNDRSHGRPP-----ASLAGSAMGLLFSFCIGLILVGRE 816  
 Db 751 PLRPGG-----SAPRRPITSSVHAKQTLATVINGIAPSEFMCVLWLFMALYR 798  
 QY 817 MKRKRKKKAELEMYAEGHNSGSDRTANNNTMKTGVKALSINLAEPKPRKLTFADL 876  
 Db 799 VRK-VQKKEIKRKKYTESLPTSG-----SCSWKLSVPEPLSINAVTEPKRKLTFAPHL 852  
 QY 877 LOATNGFHNDSLIGSGFGDYVYKALIKDSNVAIKLHVSGCGREMAEETIGKIKH 936  
 Db 853 LEATNGFSAETMVGSGFGDYVYKALIKDSNVAIKLHVSGCGREMAEETIGKIKH 912  
 QY 937 RNLVPLLYGCKVDERLVLNVEWKYSLDVLQD--PKKGVYKLISTRKAIASAGL 994  
 Db 913 RNLVPLLYGCKVDERLVLNVEWKYSLDVLQD--PKKGVYKLISTRKAIASAGL 972  
 QY 995 AFLHNHCSPIIHRDKSSNVLDENLEKRVSDFGMARISAMDTLHVSSTLAGTPGYVP 1054  
 Db 973 AFLHNHCSPIIHRDKSSNVLDENLEKRVSDFGMARISAMDTLHVSSTLAGTPGYVP 1032  
 QY 1055 PEYYSQFRCSTKGDVYSYGVVLELITGKRPDSDFG--DNNLVGVWYQ-HAKLRISVVF 1112  
 Db 1033 PEYYSQFRCSTKGDVYSYGVVLELITGKRPDSDFG--DNNLVGVWYQ-HAKLRISVVF 1092  
 QY 1113 DPELMKEDPALIEILLQHLKVAACLDPRAMRPTMVMAMFKFIOAGSGIDSGSTIRS 1172  
 Db 1093 DPELMKEDPALIEILLQHLKVAACLDPRAMRPTMVMAMFKFIOAGSGIDSGSTIRS 1145  
 QY 1173 IBDGGSTIEMDMSIKETP 1192  
 Db 1146 -----SLDEFSLKETP 1156

RA Salzberg S.L., Fraser C.M., Venter J.C.;  
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 thaliana";  
 RL Nature 402:761-768 (1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Lin X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RP [3]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Bani J., Chan M.M., Chang C.H., Chang E., Dale J.M.,  
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,  
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,  
 RA Gernin J.P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
 RA Kamaya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
 RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,  
 RA Sakurai T., Satou M., Seki M., Shim P., Southwick A., Shinozaki K.,  
 RA Davis R.W., Ecker J.R., Theologis A.;  
 RT "Full length cDNA of gene At2g01950 (GI:15226381).";  
 RT Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 CC 1-SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL; AC006532; AAD20088.1; -  
 DR EMBL; AY074313; AAL67010.1; -  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR001611; IRR.  
 DR InterPro; IPR003592; IRR\_out.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR InterPro; IPR004040; STY\_pkinase.  
 DR Pfam; PF00560; IRR; 21.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR00019; LEURICRPT.  
 DR Prodom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00370; IRR; 18.  
 DR SMART; SM00221; STYK; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.  
 DR PROSITE; PS00106; PROTEIN KINASE ST; 1.  
 KW ATP-binding; Kinase; Receptor; Serine/threonine-protein kinase;  
 KW TRANSFERASE.  
 SQ SEQUENCE 1143 AA; 125675 MW; 7D1C88493F27A94E CRC64;  
 Query Match 40.6%; Score 2509.5; DB 10; Length 1143;  
 Best Local Similarity 46.2%; Pred. No. 2.8e-155;  
 Matches 556; Conservative 183; Mismatches 353; Indels 111; Gaps 25;  
 QY 9 LSVTTLFFPFFSLSFOASPSQSLYREIHQLISFKVLPD--KULLPDWSSNKNPCTPDG 66  
 Db 14 IQISFIFLTLHSQS-SSSDQSSILKTDLSLSFKTMIQDDPNMILSNNSPRKSPQFSG 72  
 QY 67 VTCRDQKYSIDLSKPLN--VGFSAVS-----SLSLTGLBSLFSNHSINGVSQFPC 120  
 Db 73 VTCIGARTEINLSGSGISGIVSFNAFTSLDSLVKLS--ENFVLN-----STSLILL 125  
 QY 121 SASLTSIDLNSLSGVTTLTSLGSCSGKLFNVASNTL---DPEKYSQGLKINS--L 175  
 Db 126 PLTLTLELSSGLIG-----TLPENFESKYNLSITLSITLVNFTGKLPDLFLSSK 179  
 QY 176 EVLDSNLSISGANVWVWVSDGCGELKHLAISGNKISGDVDVSRVNLFLDVSSNFS 235  
 Db 180 QTLDSLNNITG-----PISGLTIP----- 199  
 QY 236 TGIFPLDSCALQHLIDISGNKLSGDFSAISTETELKLNISNNOVGPPIPL--PIKSL 293  
 Db 200 -----LSSCVSMYLDPSGNSISGYISDSINTNLKSNLSYNNFDGIPSPFELKIL 254  
 QY 294 QYLSLAENKKTGTGIPPLSGACDTLTGLDLSGNHFYGAVPPEFGSCSLNESIALSNFS 353  
 Db 255 QSLDSLNRRLTGMIPPEIGDTCSLQNLRLSYNNFYGIVPESLSSCSMWLQSLDLSNNIS 314  
 QY 354 GELPMDTLKMRGKVLVDLSPNFSGELPESLTLNSA--SLTLDSNNSNPGPLPLNLC 411  
 Db 315 GPFWTLIRSFGLQILILLNNLISGDFP--TSISACKSLRIADFSNNSNPSGVIPLDLC 371

QY 412 QNPXNTLOEYIYLNNGFTGKIPPTLSNCEIYVLSHFNYSGLTIPSSGSKLRDLXL 471  
 Db 372 PGAA-SLEBRRLPDNLVGTGIPPAISQCSLRITDLSLNYLNGTIPPELGNOKLEQFPA 430  
 QY 472 WLMMLGEIPEOIMAYVKTETLLIDFNDLTGELIPSGISNCTNIMWISLNNRIYGEIPK 531  
 Db 431 WYNNINAGEIPPELIGKQNLKDLILNNQNLGELIPPEFNCNSINEMWFSNRLTGEVPRD 490  
 QY 532 IGRLENTAIKLSNNSFGNIPPELGDRLITLMDNTLNFCTIPAAAFKSG-KIANN 590  
 Db 491 FGLISRLIAVQILQNNFTGIPPELGCCTLLVWLDNTNLTGELIPRIKRGOPSALLSG 550  
 QY 591 FIKGKRVVYIKNDGKKECHGAGNLEFQIRSEQLNRSLSTRNPNCTISRYGHTSPTE 650  
 Db 551 LLSGNTMAFPAVNG--NSCGVGVGVFSGIRPERLLQIRLSGCPFT-RMVGPIISLF 607  
 QY 651 DNNGSMFLDMSYNNLSGYIPKEIGMPYLPILNHNIDSGSIPDEVDLRLGNTLIDS 710  
 Db 608 TRYQITREYLDLSYQLRGKIPDEIGEMIALQVLELSHNOISGELPFTTIGQKNLGVFDS 667  
 QY 711 SNKLDGRIPOAMSLIMLTLEIDLSNNLSGPIPEMGQFETFPFAKPLNNGLGCGYLPFC 770  
 Db 668 DNRLOGOIPESFSLVQIDLSNMLTGPIDRGOLSTLPATQYANNPGLGVPLPEC 727  
 QY 771 DPNAD--GYAHHQRS-HGRPPASLAGSVAMGLIFSVCIFGLIIVGEMRRKRRKEA 826  
 Db 728 KNGNNOPLPAGTEBGRKAKHGTRASWANSIVLGVLSAASVCILIIWAIIVARRARDAD 787  
 QY 827 ELEMYABGHGNSGDRNTANNWKLTGYEALSITMLAEPKPRRLTFRADILQATNGFND 886  
 Db 788 AKMLHSLOAVNSA-----TTWKIEKEKEPLISIVATFQRLKFKSOLLEATNGSAA 841  
 QY 887 SLISGGGFDVYKAILKDSAVAIKCLIHVSQGDREEMMETITGKIKHNLVPLLYGC 946  
 Db 842 SMIGHGGGEPKATLKQSSVAIKLIRLSCQDRBEMAMETLTKIKHNLVPLLYGC 901  
 QY 947 KVGDERLLVNEWKYSGLEDVLODPKKGVR--LKLSTRKIAIAGSARGLAFHNNSPH 1004  
 Db 902 KIGERLLVYEMQYSGLEEVYHGPRTGCKERRILGWBERKKIAGAAGKLCFHHNCIPH 961  
 QY 1005 ITHRMKSNVILIDENLARYSDPGMARLMSMDTHLSVSTLAGPGVPEYQSFRC 1064  
 Db 962 ITHRMKSNVILIDENLARYSDPGMARLMSMDTHLSVSTLAGPGVPEYQSFRC 1021  
 QY 1065 TKGDVYSGVVLLELTGRPTSDPGDNNLVGVVYKQAKL-RISDVPELMKEDPAL 1123  
 Db 1022 AKGDVYSGVVLLELTGRPTSDPGDNNLVGVVYKQAKL-RISDVPELMKEDPAL 1081  
 QY 1124 EI-----ELLOHLYAVACLDPRAMRRPTVWQVAMKEIOAGSGIDSGSTI 1170  
 Db 1082 SLNKEGPEGVIVKEMRLYLALRCVDDPSPKRPNNLQVAVASLRELR-GESENNSHS 1140  
 QY 1171 RSI 1173  
 Db 1141 NSL 1143

RESULT 8  
 094LN2 PRELIMINARY: PRT: 1110 AA.  
 AC 094LN2;  
 DT 01-DEC-2001 (TREMELREL. 19, Created)  
 DT 01-DEC-2001 (TREMELREL. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMELREL. 21, Last annotation update)  
 DR Putative receptor protein kinase.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OC NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Wing R.A., Fritsch D., Presting G., Wood T., Yu Y., Soderlund C.,  
 RA Kim H., Rambo T., Henry D., Simmons U.,  
 RT "Rice Genomic Sequence."  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC078891; AAK52544.1;  
 DR InterPro: IPR000719; BUK\_pkinase.  
 DR InterPro: IPR001611; LRR.  
 DR InterPro: IPR000504; RNA\_rec\_mot.  
 DR InterPro: IPR002290; Ser\_cyt\_pkinase.  
 DR Pfam: PF00560; LRR\_19.  
 DR Pfam: PF00069; pkinase; 1.  
 DR ProDom: PD000001; BUK\_pkinase; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_APP; UNKNOWN\_1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00030; RR\_RNF\_1; UNKNOWN\_1.  
 DR ATP-binding; Kinase; Receptor; Transferase.  
 KW SEQUENCE 1110 AA; 118109 MW; B6723380C0A8E94 CRC64;

Query Match 36.2%; Score 2240; DB 10; Length 1110;  
 Best Local Similarity 44.2%; Pred. No. 1, 1e-137;  
 Matches 510; Conservative 172; Mismatches 361; Indels 112; Gaps 30;

QY 39 LISFDVY--PDKNLLPDWSSNNKPCTFDGYTCR-DKVTSIDLSKPLAVNGEAVSSSL 95  
 Db 28 LIRFAFVHXDPGRVLSM--VDPGCRWRTVCNGDRVTELDAA-----GGLAGREL 81  
 QY 96 LSLTLESLFSLNSHINSV---SG--FKCASLTSLDLSNLSGPTVTLTSGSCGL 150  
 Db 82 AALSGDLTLCRLNLSGNEHIVHDGLVLPALLQDLDSGGIAGRLP----- 131  
 QY 151 KPLAVSNVTLDPFGKVSQGLKINSLEVLDSANSISGANVGVVLSQCGELKHLAISGN 210  
 Db 132 GFLACYNLTD-----VSLARNMLTG-ELPGMLA--SNIRSFVSGN 171  
 QY 211 KISGVD--VSRVNLFEFDVSSNNFSTGI-PFLGDCSALQHLDISGNLSGDFSRALSTC 268  
 Db 172 NMSGDISGVSIPATLAVIDLSGNRFTGAIIPSLSGCAQLTTLNLSYNGLAAGIPEGICAI 231  
 QY 269 TELKLNLSNNGFVPIPP---LPLKSLQVLSLAENFTGEIPDFLSGACDTLTGDLDS 324  
 Db 232 AGLEVLDVSWNLHGTALPPGLGRNACASLRVLRVSSNISGISPELS-SCGALRLDVA 290  
 QY 325 GNFHYGAVP-PFPGSCSLSSALSSNNPSEGLPMDTLKRGKLVLDSPREFGELPE 383  
 Db 291 NNVVSGGIPAAVIGNLTAVESILLNNFISGSLP-DTIAHCNLEVADLSKISGALP- 348  
 QY 384 SLTNLSASLITLDSNNSNGPILPNIQNPKNITLOEYIYLNNGFTGKIPPTLSNCELY 443  
 Db 349 -----AELC-SGPAALBELRLPNNLVAGTIPPLGSLNCSRLR 383  
 QY 444 SLHLSFNYSGLTIPSSIGSLSKLRDLKMLNMLEGRIPEOLMYVKTETLLIDFNDLTGE 503  
 Db 384 VIPDSINYLGRIPPELIGRLALREKLVMMFNGLDGRIPADLQCCRLRLTLINNFIQGD 443  
 QY 504 IPGSLNCTNLANWISLNNRLTGEIPKMTIGRLBNALILSNNSFGNIPPELIGRSLI 563  
 Db 444 IPVELFNCCTGLEWVSLTSNOITGTIRPEGRSLRLAVIQLANNSLAGELPRELGNSSIM 503  
 QY 564 WLDNTNLTENGTLPAAMPFGSGKI-AANFLAGKRYVYIKNDGKKECHGAGNLEFQIR 622  
 Db 504 WLDNSNRLTGEIPRIKRGOLSTPLSGISGNTLAFVANG--NSCGVGVGLFPAIR 561  
 QY 623 SEQLNRLSTRNPNCTISRYGHTSPFDNNGSMFLDMSYNNLSGYIPKEIGMPYLPFI 682  
 Db 562 PERLLQVPTIKSCDFT-RIVSGAAGVGTREYQLEVLIDLSYNSLDEIPEELDMDVAVLV 620  
 QY 683 LNLGNDISGIPDEVDLGLNITLDSNKLDRIPQMSALITMLTEIDLSNNLSGPI 742  
 Db 621 LIDLARNMLTGEIPASIGRLNLLGVFDVSRRLRGQGIIPDSNSLFLVQIDISNNLSGEI 680  
 QY 743 PEMGQFETFPFAKPLNNGLGCGYLPFCDP-----SNADGYAHHQSRGRRPASLAGS 795

Db 681 PQRGQSTLPASQVAGNPGLCMPLEPCGDRLPATMSGLAAASTDPPPRAAVTWANG 740  
 Qy 796 VAMGLIFS--FVCIFGLIIVGEMRRRRRKEALEMTVAEGHSGDRANTNTMKL-TG 852  
 Db 741 VILAVLVASAGLCAAIWVAARARREVSAMMLSLD-----GTRTA--TWKLGKA 793  
 Qy 853 VVEALISINLAEEKPKLTPADLLQATNGFINDSLISGGGGDYKAILKDGSAVAIKK 912  
 Db 794 EKEALISINATVQROKRLKTLTQLEATNGFSTASLISGGGGEVFKATLKDGSCVAIKK 853  
 Qy 913 LIHVSQGGQREEMAEEMETIGIKIKENLVPLLGCKYGDRELLVNEVMKXSLIEDYLQDP- 971  
 Db 854 LIHLSVQGGQREEMAEEMETIGIKIKENLVPLLGCKYGDRELLVNEVMKXSLIEDYLQDP- 913  
 Qy 972 -KKGVKCLKLSTRRKIAIGSAGLAFLHNCSPHIIHDMKSSNVLLDENTLAEVSDPGM 1030  
 Db 914 GKSASPAWSMEQRKIVARGAARGLCFLHNCIPHIIHDMKSSNVLLDQDMEARVADPGM 973  
 Qy 1031 ARLMSAMDTLHLSVSTLAGPGVPPREYVOSFRCTKGDVYSYGVVLLBLTGKRTDSDP 1090  
 Db 974 ARLISLMDTHLSVSTLAGPGVPPREYVOSFRCTKGDVYSYGVVLLBLTGKRTDSDP 1033  
 Qy 1091 PGDNVLGVNVOHAKRLISD-----VFPDELKEDPALEIELQHLKVAVACLDPRAMR 1145  
 Db 1034 PGDTNLVGVN-----KMKYGDGAGKEVLDELVEGADAD-EMARMDMALQCVDDPFSKR 1088  
 Qy 1146 PPMVQVAMFEKIQ 1160  
 Db 1089 PNMQLQVAMRELDA 1103  
 RESULT 9  
 Q9LYN8 PRELIMINARY; PRT, 1192 AA.  
 AC Q9LYN8;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Receptor-like protein kinase-like protein.  
 GN T28014.220.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 CC Eucosida II; Brassicales; Brassicaceae; Arabidopsids.  
 CX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,  
 F. Rud S., Lemcke K., Meyer K.F.X.;  
 F. Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE SRR/THR FAMILY OF PROTEIN KINASES.  
 DE EMBL, AL163662, CAB87284.1;  
 DE InterPro: IPR000719; Euk\_kinase.  
 DE InterPro: IPR001611; LRR.  
 DE InterPro: IPR003592; LRR out.  
 DE InterPro: IPR002290; Ser thr kinase.  
 DE InterPro: IPR004040; SHY\_kinase.  
 DE InterPro: IPR001245; Tyr\_kinase.  
 DE Pfam: PF00560; LRR; 25.  
 DE Pfam: PF00069; pkinase; 1.  
 DE PRINTS: PR00019; LRR/LCHRP.  
 DE PRINTS: PR00109; TYRKINASE.  
 DE ProDom: PD000001; Euk\_kinase; 1.  
 DR SMART: SMO0370; LRR, 25.  
 DR SMART: SMO0221; STYK; 1.  
 DR PROSITE: PS00107; PROTEIN KINASE ATP, 1.  
 DR PROSITE: PS00011; PROTEIN KINASE DOM, 1.  
 DR PROSITE: PS00108; PROTEIN KINASE ST, 1.  
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.

SQ SEQUENCE 1192 AA; 12979 MW; D02C8B8FE6B09F7D CRC64;  
 Query Match 26.0%; Score 1608.5; DB 10; Length 1192;  
 Best local Similarity 35.2%; Pred. No. 2.4e-96;  
 Matches 453; Conservative 177; Mismatches 421; Indels 235; Gaps 36;  
 Qy 11 VTTLPFFSFFSLSPQASPSQSLYREIHOLISFEDVLPDKNLDPW--SSNKNPCTPDGYT 68  
 Db 4 LTFALFLPFLF--SFSSSAIVDLSEFTSLISFGRSLLENPSLLSSNMVSSASHCDWVGT 61  
 Qy 69 CRDDKXTSIDLSKPLNAGVSAVSSLSITGLSESLFNSHINSVS---GFKCSASL 124  
 Db 62 CLDGRVNSLSLPLSLR--GQIPKEISLKLRELCLAGNPFSGKIPPEIWMLK---HL 115  
 Qy 125 TSIDLSENLSGCVTTTLTSLGSCGKFLNVSNTLDPPGKV--SGGLNLSLEVDLSA 182  
 Db 116 QTDLDGNSLITGLPLR--LSELPLLYLDLSDN--HFGSLPPSFSLPALSLSDVN 171  
 Qy 183 NSISGANVGVWLSDGCGELKHLA--ISGNKISGDV----- 216  
 Db 172 NSLSGE-----IPPEIGKLSNLSNLYMGLNPSGQIPSEIGNISLLKNFAAPSCFNGP 225  
 Qy 217 ---DVRGVNLEFLVDVSSNNPSTGT-----PFLGDCSALQ 248  
 Db 226 LPKEISKRLHLADLDLSTYPLKCSIPKSGELNLSILNLVSAELGLIPPEIGNCKSLK 285  
 Qy 249 HLDSGNKLSGDFSRALSTCTELKLNIS--NPFVGPPL--PLKSLQVLSLENKRT 304  
 Db 286 SLMSFNSLSGLPLPLSL--ELPLTFSAERNQLSGSLPSMKGKVKVLDLSLLANRPS 342  
 Qy 305 GEIP-----DFLSGA-----CDTLGLDLSGNHFGYAVPPFGSCSL 341  
 Db 343 GEIPHEIEDCPMLKHLSLANLSLGSIPRELCSGSLBAIDLSGNLSGTIEVDDGSS 402  
 Qy 342 LESLASSNNPFGELPMDTLKKRGKXVLDLSPNPSGSLPSL---TNL----- 388  
 Db 403 LGELLTNNQINGSTPED-LMKL-PLMALDLDSSNFTGRIPLSLKSTNLMFTASYNRL 460  
 Qy 389 -----SASLTLTLDSSNNPFGILPNCQNPKNLTGLYLNQNGFTKIPPTLSN 438  
 Db 461 EGYLPALFIGNAASLKLVLSDNQLTGEIREF--GKLSLSLTNANANPQCKIPEVLD 518  
 Qy 439 GSELVSLHSPNYLSTGTPSSLSGLKLDLKLMLMLEGEIPQ-----ELMY 486  
 Db 519 CTSLTTLTLDGNNLQGIIPDKITALAQCLVLSYNNLSGSLPSKPSAVFHQIEMPDLSF 578  
 Qy 487 VKTLFLLIDPNDLGEIRISGSGNCTNLMWISLNNRLTGERPKMIGLENTAIIKLSNN 546  
 Db 579 LQHHGIFDLSTYRSLSGPIPEELGECIVLVEISLNNHLSGELPASLSRLTNLTIDLSGN 638  
 Qy 547 SFGNIPDELGCPSLIMDLNTLNFNGTIPAMAFKQSGKLANFLAGRYYYINKDMK 606  
 Db 639 ALTGSLPKMGSKLQGLANTLANNQNGHIP----- 669  
 Qy 607 KECHGAGNLLFQGISQOLNLSLRNPNCTISRYGGHTSPTPDNNGSMFELDSYML 666  
 Db 670 -----ESFGLGSLVTKNLTKNLDGPVPASLGNLKLTLTMDLSFNNL 712  
 Qy 667 SGYIPKEISMPYFLTNLGHNDISGTPDEVDLGRNLIDLSNNKLDGRIPQAMSLT 726  
 Db 713 SGLSSELSTMEKLVGLYIEQNKFTGEIPSELGNLTQLEVDVSENLSLGEIPTKICGLP 772  
 Qy 727 MLTEIDLNNNLSGPIPEWGCQETFPFAKFLNPNGLCGVPLPCDPSNADGVAHQRSHG 786  
 Db 773 NLEFLNLAKNNLRGVPSGVQDPSKLLSNKELCG---RVGSGC-----KLEG 821  
 Qy 787 RRPASLAGSVAMGLIFS--FVCIFGL--IIVGEMRR--RRKKEALEMTVAEG--HG 836  
 Db 822 TKLSAMGIAGLMGLFTIIIVFVPSLRWMAKTRVKQRDDPERMEBSRLGFDQNLVYF 881  
 Qy 837 NSGDPNTANTNTMKLVGEALISINLAEEKPKLTPADLLQATNGFINDSLISGGGG 896  
 Db 882 LSGSRS-----RPPISINIMFEQPLKVLKRLGDIIVEATDHFESKNITIDGGGRT 930

QY 897 VYKALDGSAAVAIKKLIHVSQGDREFFAMMETIGIKERNVLPLIGYKVGDELLIV 936  
 DB 931 VYKACLPEKTVAAVKKLSBAKTQGNREFFAMMETIGIKERNVLPLIGYKVSFSEKLLIV 930  
 QY 957 EVMKYSJEDVLQDPKKGIVLKLSTRKIAISARGIAFLHNHCSPHIHRDKSSNVL 1016  
 DB 991 EYMWGSLDHMLRNQTMLEVLWDSKRIKIAVGARGLAFHGHGPTPIHTRDKASNVL 1050  
 QY 1017 LDENLEAVSDPGMARLMSAMDTHLSVTLTACPGVPEYQSFSGSTKGDVYSYGV 1076  
 DB 1051 LDGDEPFPVADFGARLARSACESHS -TVIAGTFCYIPPEYQSGARATTKGDVYSFVIL 1109  
 QY 1077 LELLTGKRPDSDFGDN--NLVGVKQHAFL-RISVPEPELMKEDPALEILLQHLK 1132  
 DB 1110 LELVYGEKPT-GPDEKSEGGNLVGMALQKINQKAVVIDPPLV--SVALKNSQLRLQ 1166  
 QY 1133 VAVAGLDDRAMRPTMVOVMAFMFKI 1158  
 DB 1167 IAMLCLAEPTAKRPNMLDVALKALKEI 1192

RESULT 10  
 Q8RZV7 PRELIMINARY; PRT; 1294 AA.

AC Q8RZV7  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative receptor-like protein.  
 GN P0413C03.19.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzoideae; Oryza.  
 OC NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
 clone: P0413C03."  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF003451; BAB86144.1; --  
 KW Receptor.  
 SQ SEQUENCE 1294 AA; 141318 MW; 88658384859822A1 CRC64;

Local Match 25.8%; Score 1593.5; DB 10; Length 1294;  
 Local Similarity 33.3%; Pred. No. 2,6e-95;  
 Mismatches 454; Conservative 175; Mismatches 434; Indels 299; Gaps 42;

QY 5 SSEFFLSVTLFFPSFSLSPFASPSQSLYREIHQILSPDVLDP-KNLLPDM-SNNKPC 62  
 DB 4 NSFWLPT---LVSFPIISAMAES---RDISTLFTLSDSITGKGFRLNMFDSPTPC 55  
 QY 63 TFGVTCRDKDXTSILDSKPLAVGSAVSSLSLTGLESLELSHSHNGVSGFKSA 122  
 DB 56 SWGICITGHNVAIDLSVPLAAPPPLCTIGAFQSLV-----RLNFGCGE----- 101  
 QY 123 SLTSLDSRNSLSGPVTTLSLSSCGSLKFLNVSNTLDFPGVSGGLKLNLEVLDSA 182  
 DB 102 -----SGELP--EALGNLQNLQYLDLSNNEHLPIDIS-LVNLKMLKENVLDY 146  
 QY 183 NSISGANVGVWVLSDCGELKH---LAISGNKISGV--DVSRCVNLEFLDVSSNNFTG 237  
 DB 147 NSLSGQ-----LSPAIAGLOHLTKLISWNSSISGLPDLGSLKLELLDIDKMTFNGS 200  
 QY 238 IP-FLDGCALQHLDISGNTLSDPSRAISTCTELKLNINISNQFVGPFP----- 286  
 DB 201 IPATFNLSCGLHFDASQNNLTGSIPTGITSITNLTLTLDSSNSFGTLPREIGOLENIE 260  
 QY 287 -----PLP-----LKSLOYLSLAENKFTGEIPPLSGACDTLTGLDLSGNHFGY 330

DB 261 LILGKNDLTGRIPQETIGSLKQKLHLHEFCQGTGKIPLWSISG--LSSITLIDISDNNF 319  
 QY 331 AVPPFGCSLLLESLALSNNPFGELPMDTLTKMRGLKVLVDLSFNEBSGELPESUTINSA 390  
 DB 320 ELPSMWGELGNLTOLIAKNAGLSGMPKE-LGNCKKLVINLSFNAIIGPIPEEPADLEA 378  
 QY 391 -----SLTLTDLSSNNPFG--PLP----- 408  
 DB 379 IYFVEKNGKLSGRPDMIQKKAKNARSIRLQNNFSGGLPVLPIQHLSPFAESNLSGS 438  
 QY 409 ---NLQNPKNLTLOELTYQNNFTGKIPTLSNCELSVLHLSFNYSGLTIPSSIGLSLK 465  
 DB 439 IPSHICQ--ANGLSHLLHNNLTGTIDEAFGCTNLTEMLDINHIGEVGYLAEL-P 495  
 QY 466 LDDKMLNMLEGHPQELMAYKTELETLIDPNDLTGEIPGLSNCNTLWNISNNELT 525  
 DB 496 LVTLELSQKFPAGMLPAELWESKTLLETSLSNNEITGPIPSIGLSTVQLRLHDNNLE 555  
 QY 526 GEIPKWIQRLNMLALIKLNNPFGNIPDELGDCSLIWLDTNPLFNGTIPAMF--- 581  
 DB 556 GPVPSGDLRLTNLSLRGNLSGIIPLALFNCKRLATLDSYNNLTGNIPLAISHLTL 615  
 QY 582 -----KQSGKIAPNFIAKRYVYIKNDGKKECHGAGNLEFQIGRSEQLNRLSTR 632  
 DB 616 LDSLSSNQLSGSIPAEICV-----GFENEAPDSSEPLQHLGLDLSYNQLTGQ 665  
 QY 633 NPQNTISRYGGHTSPTFPNNNSMFLDSYNNLSGYI PKEISGMPYFLINTGND--- 689  
 DB 666 IPTSI-----KNCAMVWNLNQLNGLTITPELIGELTNLTSLSNNEVFG 712  
 QY 690 -----ISGIPDEVD-LRGLNLTLDLSNNKLDGRIPQANSALTM 727  
 DB 713 PMLPWSGPLVQQLGILLNNHLDGSPAKIGQILPKIATLDSNVALTGTLPOSLCNVY 772  
 QY 728 LTEIDLSNNLSGPI-----PEMQEFPFPAKFLN 758  
 DB 773 LNHLDVSNHLSGHIQFSCPDKEYSSTLLFPNSSNHFSGSLDSISNFTQLSTLIDHN 832  
 QY 759 NP-----GLCG-----YPLPR-- 769  
 DB 833 NSLTGRIPSLDLSLNLVLDLSNNLYAIPCGICNIFGLSPANFSGYIMYSLADA 892  
 QY 770 ----CDPSNADGYAHHQRSHGRPPASLAGSVAMGLLFSFVCI FGLVGRNKRER-RKK 824  
 DB 893 AGGICSTNGTDKALHPHYRVRALITCA-----FTVILIVLAVYLRRLKLVSR 945  
 QY 825 EALEMYAEGHNGSGRTYANNIYW-KLTGVK--EALSTINLAPEKPKLCTPADLQATN 881  
 DB 946 PLAFE-----SASKAKATVEPTSTDELGKRSREPLISINLAFETALLETADILKATE 1000  
 QY 882 GFINDSLISGGGDDYKAILKDGSAVAIKLIHVSQ--QGDREFFAMMETIGIKERNVL 939  
 DB 1001 NFSKVHILIGGGGIVYKALPEGRVAILKRL-HGHQHQGDREFFAMMETIGIKERNVL 1059  
 QY 940 VPLIGYCKVDEERLIVNEVWKYSLEDVLQDPKKGIVLKLSTRKIAISARGIAFLH 999  
 DB 1060 VPLIGYCKVDEERFLIYEWENGSLLEMLLRNADALEALGMPDLKICIGSARGIAFLH 1119  
 QY 1000 NCSPHIHRDKSSNVLLENLEARYSDGMAFLMSAMTHLSVTLTACPGVPEYQ 1059  
 DB 1120 GFVPHIHRDKSSNVLLENLEARYSDGMAFLMSAMTHLSVTLTACPGVPEYQ 1178  
 QY 1060 SFRSGTGDDYYSYGVVLELTGKRPDSDF--GDNNLGVWV--QAKKRIISVPEPELM 1117  
 DB 1179 TMSKTTGDDYYSYGVVLELTGKRPDSDF--GDNNLGVWV--QAKKRIISVPEPELM 1237  
 QY 1118 KEDPALEI---ELLQHLKVAVACLDDRAMRPTMVOVMAFMFKI 1156  
 DB 1238 ---PVSVWREQMARVLAIRADCTADDPFRPTMLEVYKGLK 1276

RESULT 11  
 Q9XEI3

ID Q9XEL3 PRELIMINARY; PRT; 1420 AA.  
 AC Q9XEL3;  
 DT 01-NOV-1999 (Tremblrel. 12, Created)  
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
 DE EST C96716(C10608) corresponds to a region of the predicted gene.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Euphorbiaceae; Oryzaeae; Oryza.  
 NC NCB1;taxid=4530;  
 RP [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, NIPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa Nipponbare (GA3) genomic DNA, chromosome 2, PAC  
 RT clone: p0437H03 (contig b).";  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DT EMBL; AP000367; BAA82393.1; -;  
 DT InterPro; IPR000719; Euk\_pkinase.  
 DT InterPro; IPR001611; LRR.  
 DT InterPro; IPR003592; LRR\_out.  
 DT InterPro; IPR002290; Ser\_thr\_kinase.  
 DT InterPro; IPR004040; STY\_kinase.  
 DT InterPro; IPR001245; Tyr\_kinase.  
 DT Pfam; PF00560; LRR; 30.  
 DT Pfam; PF00069; pkinase; 1.  
 DT PRINTS; PRO0019; LEURICHRPT.  
 DT PRINTS; PRO0109; TYRKINASE.  
 DT ProDom; PD000001; Euk\_pkinase; 1.  
 DT SMART; SM00370; LRR; 24.  
 DT SMART; SM00221; STYK; 1.  
 DT PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DT PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
 DT PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DT ATP-binding; Serine/threonine-protein kinase; Transferase.  
 SQ SQUENCE 1420 AA; 153401 MW; B938CE5B9AE92F5 CRC64;  
 Query Match 24.3%; Score 1503; DB 10; Length 1420;  
 Best Local Similarity 32.0%; Pred. No. 2,5e-89;  
 Matches 439; Conservative 186; Mismatches 462; Indels 284; Gaps 40;  
 QY 6 SFPLSVT-ILFPSPFSLFQASPSQSLYR-RIHOLISF-KOVLDDKMLDPMWSNKK-P 61  
 DB 99 SFIVPVSRAVYHCVLTLLCFIPITALLAESDITKULFAIKAIYAGKPELHWELTPP 158  
 QY 62 CTPEGVTCDDDKTSTIDLSKPLANGFSVSSLLSTGLSELFLSNHINGSVGFCS 121  
 DB 159 CMMSSGISCVGLTVVAILDLSSTPLVDF--PQIITAFOSLVRLINVSQCGSGELPEAMVN 215  
 QY 122 AS-LTSLDLSRNSLSGPVTVTLTSLSCSGKFLKLVNSNTLDPPKVSQGL-KLNSLEYLD 179  
 DB 216 LQHQLHLDLSNDQGLGFLP--ASLFDLKLKLVNLDNM--FSQQLSPALNHLQQLTVLS 271  
 QY 180 LANSISGANVGVWVLSQCGELKHLAISGNKISGDVVSRCVNLFEFVDSNNFSGIP 239  
 DB 272 ISTNSFSGG-----LPELIGSLK-----NLEIYDITNAFSSGIP 306  
 QY 240 -FLGPCSLMHLDISGNLSDGFSALSTCTELKILNINSQVGFIP--LPLKSLQYL 296  
 DB 307 ASFSLVSLRLIYDANNVLTGSIIPGIRALVNLVLDLSSNGVIGALPKELCQKLNQSL 366  
 QY 297 SLAENKFTGEIPDFLS-----GACDVLTLGLDLSGNH----- 327  
 DB 367 ILSDNELTGISPEEIGNKOLEVNLKCNLMQVPISTIGLLELLELYISFNSFSGELP 426  
 QY 328 -----FYGVPPFSGSCSLLESALSSNNFSGELPMDTLKMGELKY 369  
 DB 427 ASVGEIHLRLQMLAKSAGFTSIPKEIKNCCKLTTVLISGNFTGIRP-BELADIVAVVL 485  
 QY 370 LLSFNEFSGELPE----- 383

DB 486 PDVEGNRLSGHIIPDIQWNSVSSISLAQMFDPGLPGLHLVSPSASBNRLSGSIPAK 545  
 QY 384 -----SLTNIS-----ASLTLFIDS 398  
 DB 546 ICGFTLOILRLANDNNLTGSIIDETFKCKKLTSLDNLHGEIPEYIALPLVSLDLS 605  
 QY 399 SNNFSGPLPLCONPKNTQELYLQNGGTGKIPTLSNCELSVLSHFNYSGLTIPS 458  
 DB 606 HNNFTGIMPDLMS--STLIDSLSDNQLTGMTESIKGLSLQSLSDRYVLQGLPR 663  
 QY 459 SLGSLSKLRDLKMLNMLEGRIPOELMYVYKTELTFLFDENDLTGERPSGLSCTUNMIS 518  
 DB 664 SIGALRLVTLTSLSGNLSLSDIPIQLFNCRNLVTLTSCNNALIGHIPALSHLTLYKNTLV 723  
 QY 519 LSNRLTGERP-----KWIGLENTAILKLSNNSFGNIPDELGDCRSLMWD 566  
 DB 724 LSRRLSGALPSELVAFSRESHSELEVOHIGLIDLSRNLRTGHIIPRAINNCSTLVELH 783  
 QY 567 LNTVLFNGTIPAMFKQSG--KIANFLAKRYVYIKNDGKKEGAGNLEFGQIR 622  
 DB 784 LQDNLSTGTPVLAELRNITITDSSNALVGPVLP-----PVPLASLQGLL 831  
 QY 623 SEQNLRLSTRNPNITSRVYGGHTSPTPDNNGMMFLMSYMLSGYIPKEI----- 674  
 DB 832 LSN-NRLSGSIPSGI-----GNILP-----QIMLDLSGNALTGITPLDLCKESLNH 878  
 QY 675 -----GSMFY-----LFILNLGHNDISGIPDEVGLNGLNLDLSNKL 714  
 DB 879 LDVSDNNISGQIPFSCHEDEKESPIPLIFPNASNNFSGSLDESISNFKLYLTLNHNLSL 938  
 QY 715 DGRIPQMSALTMLTEIDLSNNLSGPIP--EMGQFE--TEPPAFINNPGLCGPLPRCD 771  
 DB 939 TGRPLAIRAVTSIYVLDLSSNDPSGTTIPCGICMGFLTP--ANFSGRDGTFPLDCA 996  
 QY 772 PSNADYVHHQ--RSHGRPPASLAGVAMGLLFVFCIFGLILVGRKKRRKKEALE 829  
 DB 997 AEEGVCANRVRDKPDPHFVLEATICIAITAVILVVLVILVYLRKKMLRRQFV 1056  
 QY 830 MYAEGHNSGDFNTANNMKLTG-----VKALSLINLAEPKPLKLTFLADLQATNGFH 884  
 DB 1057 LVPAGDNMADEHTTLLSN-NLIGRRMRKRRPSPINLATEFHAVRVTVVEIMRATNGFD 1115  
 QY 885 NDSLIGSGGFGDVYKAILKDSAAVAILKLIHVSQ-----GDREPMAMETIGIKKHNLV 940  
 DB 1116 GNVAVVDGGFVTVYRAELPGRRAVAVKRL-HGVGRKPOGGEREFRAEMETGVKVRHNLV 1174  
 QY 941 PLLGVCXGDERLLVNEVWKGSLEEDVLQDPKKGVLKLTSTRKIALISARGIAFLHNH 1000  
 DB 1175 PLLGVCAGDERFLVYEMHESLEEDLRG--GGGALGMPERLITCGAARGLAFLHNG 1232  
 QY 1001 CSPHILHRDKSSNVLDENLEAVSPDGMARLMSANDTLSTVLAGTGYVPEPYOS 1060  
 DB 1233 FVPHVIRHADVSSNVLAGEGLQPRVSPGLARITISACEYHS--TVLAGTGYVPEPYAIA 1291  
 QY 1061 FRCTGKGVYIYGVLLLELTGKAPT-----DSPDFGNNILVGMWK--QHALTRIS 1109  
 DB 1232 MKCTAKGDIVYFVGVMLELTGPRPTWSABVYTABDDEGGSGSLVGRMMARGRGG 1351  
 QY 1110 DVPEPELMKEPDAIEILLQHLKAYAVACLDPRAMRRPTVYQVAMFKEIQ 1160  
 DB 1352 EVFDACLPVSGARE--QMARVLDVARDCTADEPWRPPTVAFAVRVGAIEA 1401  
 RESULT 12  
 Q9SN91  
 ID Q9SN91 PRELIMINARY; PRT; 1232 AA.  
 AC Q9SN91;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
 DE Leucine rich repeat-like protein.  
 OS Arabidopsis thaliana (Mouse-ear cress).



OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Terry N., Ardiles W., Buyschaert C., Dasseville R.,  
 De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,  
 Villarejo R., Gielens J., Van Montagu M., Bancroft I., Mewes H.W.,  
 Mayer K.F.X., Lemcke K., Schueller C.,  
 RA Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP EU Arabidopsis sequencing project;  
 RA Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RL [3]  
 RN SEQUENCE OF 1-305 FROM N.A.  
 RP Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.,  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RL [4]  
 RN SEQUENCE FROM N.A.  
 RA Terry N., Ardiles W., Buyschaert C., Dasseville R., De Clerck R.,  
 De Keyser A., Neyt P., Rouze P., Van Den Daele H., Villarejo R.,  
 Gielens J., Van Montagu M., Mewes H.W., Lemcke K., Mayer K.F.X.,  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RL [5]  
 RN SEQUENCE FROM N.A.  
 RP EU Arabidopsis sequencing project;  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL: AL022224; CAA18239.1; -  
 DR EMBL: AL161552; CAB79014.1; -  
 DR HSP: P08631; 1AD5.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR001611; LRR\_out.  
 DR InterPro: IPR003592; LRR\_out.  
 DR InterPro: IPR002290; Ser\_thr\_pkinase.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR Pfam: PF00560; LRR; 30.  
 DR Pfam: PF00069; pkinase; 1.  
 DR PRINTS: PRO0019; LEURICHRPT.  
 DR PRINTS: PRO0109; TYRKINASE.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SMO0370; LRR; 30.  
 DR PROSITE: PS00107; PROTEIN KINASE ATP; UNKNOWN\_1.  
 DR PROSITE: PS00101; PROTEIN KINASE DOM; 1.  
 DR PROSITE: PS00108; PROTEIN KINASE ST; 1.  
 DR ATP-binding; Serine/threonine-protein kinase; Transferase.  
 RN [6]  
 RP SEQUENCE 1232 AA; 135534 MW; EAD6322C8B967726 CRC64;  
 Query Match 23.0%; Score 1419; DB 10; Length 1232;  
 Best Local Similarity 31.2%; Pred. No. 6.2e-84;  
 Matches 402; Conservative 221; Mismatches 441; Indels 224; Gaps 41;  
 QY 9 LSVTLTFFSFSFSSFGQSPSSQSLYREIHQLSFKDVL---PDKNLLPDMS-NKRNPT 63  
 DB 4 LVLLTLFLICFSGSL---GQPG-IINNDLOTLEVKKSLVTNPQEDPLRQNSNDINVC 59  
 QY 64 PDGVTC-----RDKVTSIDLSKSPINVGSAVSSLL 96  
 DB 60 WGVTCNDNGLFRVIALNLGLGLTGISPWFGFRDNLTHLDLSSNNL-VG--PIPTALS 116  
 QY 97 SLTGESLFLSNHNGSVG-----FKCS 121  
 DB 117 NLTLSLESLFSSNQLTGEIIPQSLVNRSLRIGDNELVGIPELTGLVNLQMLALAS 176  
 QY 122 ASLT-----SLDLSRNLGSPVTLTSLGSCGKLKELNVSSNTLDPPGKYS 167  
 DB 177 CRTLGPSPQGLRVLVQSLILQDNYLGGPIP--AEIAGNSDLTVTAEMNLN--GTIP 232  
 QY 168 GGL-KINSLSEVLDLSANSGANVVGWVLSDCGE--LKHAIISGNKISGDV--DVSRIC 221  
 DB 233 AELGRLENLEIHLNANSLTGE-----IPSQGEMSOLOLYISLMANQGLIPKSLADL 286

QY 222 VNIIEFLDVSSNNPSTGIP-FLDCCSALQHDISGNKLSGDFERRI-STCTEIKLINISSN 279  
 DB 287 GNLQTLIDANNLTGRIPEEFNNMSQQLDLVIANHNLSSLPKSLCSNNNTLEQLVSGT 346  
 QY 280 QFVGPPIP--LPLKSLQVLSLENKFTGEIPDPLSGACDTLTGLDSGNHFGAVPPRG 337  
 DB 347 QLSGEIPELISKQSLKQDLDSNNLAGSIPALPELVE-LTDLVHNNTLEGTLSPTSIS 405  
 QY 338 SCGLSLSLSSNNFSGEIPMDTLKRGKLYLDSFNESGELESLNLASLITLDL 397  
 DB 406 NLNLTQWLIVYHNHLEKLPKE-ISALRLEVLFLYENFSGEIPGEINCT-SLKMIDM 463  
 QY 398 SNNFSGPILPNLQNPKNLTQELVYNNNGFTGKIPPTLSNGSEVSLHSPNYLSGTP 457  
 DB 464 FGNHGEIIPPS--GRKELNHLRQNEIVGCLPASLGNCHQNLITLDADNQLSGSIP 521  
 QY 458 SSLGSLKRLDKLWMLMGEIIPQELMYVKTLETILDPNDLTG----- 502  
 DB 522 SSFQFLGLBQMLVYNNSLQNLPSLISRLNLRINLSHNRINGTILHPLCGSSSYLSFD 581  
 QY 503 -----EIPSGLSNCTNLNMTSLSNRNLGTEIPKWLGRLEMLATIKLSNNSFGSNIPD 554  
 DB 582 VTNNGFDEIPELGNQNDLRRLGNQLTGKIPMTLGRILSLSDMSNALTGPIPL 641  
 QY 555 ELGDCRSILWDLNLTNLPNGTIPAMFK--QSGKIANTFAGKRYVYIKDGKKKCHGA 612  
 DB 642 QVILCKLTHIDINNPLSGIIPWLGKLSQGLBK---LSNQFV---ESLPTELFNC 694  
 QY 613 GNLL-----EFGIRSEQLNLRSTNPNCTISRYVGCHTSPTFDNNGSMFLDMSYNN 665  
 DB 695 TKLIVLSDGNSLNGSIPOEIGNGALVNLNDRQSSGLPQAMGKSLKYLRLSRNS 754  
 QY 666 LSGYIPKEIGSMPL-FLINGNHNDISGIPDEVGDRLGNITLDSNKL DGRIPQMSA 724  
 DB 755 LTGELPVEIGQLDQSLDLSYNNFTGDIPTGTLSKETLDSHNQLTGEVPGSVGD 814  
 QY 725 LTMLEIDLNNNNLSGPIPEMGQFETPPAKFLNPPGLCGYPLPRCPDSNADGYAHHQR 784  
 DB 815 MSLQIVLVNVSFNNLGGKLLK--QFSRWPADSFLNGTGLCGSPLSRCN----- 859  
 QY 785 HGRPAPSLAGSVAMGLFSPFCIRGLIVGEMRRRR--KKEALEMYAEHGNSGDR 841  
 DB 860 ---RVRTISALTAIEMLVIALF-----FKORHDFPK-----VHGSTATYT 899  
 QY 842 TANNNTWKLTVKEALSNLAFAEKLK-----LTFADLLQATNGFHNDSLISGCF 894  
 DB 900 SSSSSS-----QATHKPLFRNAGSKSDIRMEDIMEATHMLSEEFMISGGS 945  
 QY 895 GDVYKAILKQSAVAIKLIVSG--QGDREPAEMETIGKIKHRVLYPLGVCYKGD-- 951  
 DB 946 GKYYKALENGEFTVAVKTLIMDDLSNKSFSREYKVTGRIHRHLVKLMGVCSSKSG 1005  
 QY 952 RLLVNEVMKYSLEDTLQDPK---KGVYKLSLSTRKIAISAGLAFLHNCSPIIIT 1007  
 DB 1006 NLLIYEMKNGSITWDLHEDKYLEKKKLLDMEARLRAVAGLAQVETLHDDCVPIYH 1065  
 QY 1008 RDKSSNVLLIDENLTARVSDFGMARLMSA-MDTHLSVT-LAGTGYVPEPYGFRCS 1065  
 DB 1066 RDKSSNVLLDSMWEALHDFGLAKVLTENCNTDNTSNWFCSGYTAPEAYSLKATE 1125  
 QY 1066 KGDVYGVVLLLELTKGRPTSPDFGDNLLVGMVYQKHKLKS---DVPDELKKEPRA 1122  
 DB 1126 KSDVYSWGIWLVIGKPTSDVFGAEDQWRVWETHLEVAGSARDKLDIKLKLPLDF 1185  
 QY 1123 LLETILOHKAVALCDDBRAMRPTWQ 1150  
 DB 1186 EEDACQVLEIALQCTKTSPOERSSRQ 1213  
 RESULT 13  
 Q9FN37 PRELIMINARY; PRT; 1036 AA.  
 ID Q9FN37

AC 09FN37;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Receptor protein kinase-like protein (Putative receptor protein kinase).  
 GN ATG53890.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLUMBIA;  
 RX MEDLINE=98162728; PubMed=9501997;  
 RA Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N., Tabata S.;  
 RA "Structural analysis of Arabidopsis thaliana chromosome 5. III. Sequence features of the regions of 1,191,918 bp covered by seventeen physically assigned P1 clones.";  
 RA DNA Res. 4:401-414(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Banu J., Banno F., Chang E., Dale J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C., Yamanura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R., Hayaishi T., Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koese E., Lam B., Lin J., Meyers M.C., Miranda M., Nishimura M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shimizu P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R., Theologis A.;  
 RA "Full length cDNA of gene At5g53890 (GI:15238872)";  
 RA Submitted (NOV-2001) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Banu J., Chan M.M., Chang C.H., Chang E., Dale J.M., Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C., Yamanura Y., Yu G., Bowser L., Carninci P., Chen H., Cheuk R., Hayaishi T., Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koese E., Lam B., Lin J., Meyers M.C., Miranda M., Nishimura M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shimizu P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R., Theologis A.;  
 RA "Arabidopsis Open Reading Frame (ORF) Clones";  
 RA Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AB007644; BAB10719.1;  
 DR EMBL; AY091180; AAL14119.1;  
 DR EMBL; AY091180; AAL14119.1;  
 DR InterPro; IPR000719; Euk PKinase.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR003592; LRR out.  
 DR InterPro; IPR003591; LRR typ.  
 DR InterPro; IPR002290; Ser\_thr\_kinase.  
 DR InterPro; IPR001245; Tyr\_kinase.  
 DR Pfam; PF00560; LRR\_19.  
 DR Pfam; PF00069; pkinase.1.  
 DR PRINTS; PR00019; LEURICHRPT.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk PKinase.1.  
 DR SMART; SM00370; LRR\_14.  
 DR SMART; SM00369; LRR\_TYR\_9.  
 DR SMART; SM00220; S\_TKc\_1.  
 DR SMART; SM00219; TYKc\_1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; UNKNOWN\_1.  
 DR PROSITE; PS00101; PROTEIN KINASE DOM; 1.  
 DR PROSITE; PS00108; PROTEIN KINASE ST; 1.  
 KW ATP-binding; kinase; Receptor; transferase.  
 SQ SEQUENCE 1036 AA; 114339 MW; 6DP9511FC2A2E261 CRC64;

Query Match 22.7%; Score 1402.5; DB 10; Length 1036;  
 Best Local Similarity 33.4%; Pred. No. 5.8e-83;  
 Matches 394; Conservative 182; Mismatches 416; Indels 189; Gaps 33;

QY 11 VTTFFSFFSLSQSPQSLVREIHQILSPQVLPDKNLLPDMSSNNKPCPTFGSVTCR 70  
 Db 4 ILLVFFVSSVSQPCPRN-----DLALRELALGALKNSVTSWLSNGRCCEMDVFCF 58  
 QY 71 ----DKVTSIDLSKPLNLFSAVSSSLISLTGLSLFSLNSHINGSVSGFCASALTS 126  
 Db 59 GSDVSGRTVTLVPEKLE---SVISKSLGELT-----ELRV 92  
 QY 127 LDRNSLSGPTVTLTSLSGCSGLKFLNLSNTLDPGKVSGLKLNLSLEVIDLSANSTIS 186  
 Db 93 LDRNSLQKEV-----PABIS---KLQQLQVLDLSHNLIS 125  
 QY 187 GANYVAVLSDGCGELKHLAISGNKISGDV-DVSRCVNLEPFDVSSNNSTGCI-PELDDC 244  
 Db 126 G-SVLGVV--SGKLIGSLNLSNSLSGLSDVGVPPGLVNLVNSNLEFGEIHEL--C 180  
 QY 245 SA--LQHLDISGNKLSGDPBSRAISTCTELKULNLSNQFVGPFP--LPKSLQYLSLA 299  
 Db 181 SSGGGLQVLDLSMNRVLGNLDGLNCSKSIQQLHIDSNRLTQGLDYLSIRELQSLIS 240  
 QY 300 ENKFTGEIPDPLSGADTLTGLD---LGNHRYGAVPPFGSCSLSESLAISNNFSGEL 356  
 Db 241 GNYLSGELSNLS---NLGKLSLISENRFSDVI PDVFGTLTLEHLDVSSNKFSGRF 296  
 QY 357 PMDTLLKMGKLVLDLSEFNEFSGELPESTYTNLSALTLTLDSSNNFSGPIIPNLCQPKN 416  
 Db 297 P-PSLSCGSKRLVLDLRNLSLGSININFTGT-DLCYDLASNFSGLPDSLGHCPC- 353  
 QY 417 TLQELVQNNNGFTGKLPPLNSCSLVSLSLHSEFNYSGLTIPSSLSGLSKLDLKLIMNML 476  
 Db 354 -MKLISLANERFGKIPDTFKLQSLFLSLNN-----SPFDSEETMANVL 398  
 QY 477 EGEIPQELVYVTLLETLIDPNDLGEIPSGLSNCTNLSLNNRLTGEIPWIGRL 536  
 Db 399 Q-----HCNLSLTLISKNFGEIEIPNVVGTGPDNLAIALNCGIRGQIPSMILNCK 450  
 QY 537 NLAILKSNNSPSGNIPELPGCSLWLDLNTNLFNGTIPPAFPGKSGKTAANFNGKR 596  
 Db 451 KLEVIDLSNNHRYGTPHMGIMBSLFTIDSNNTLIGALPVALTEKTLRLNGTASQM 510  
 QY 597 Y-----YIKNDKKKECHGAGNLEFQIRSEQLNRLSTRNFCNTSR---VYGGHT 646  
 Db 511 TDSGIPLYVYK-----NKSNGLPYNGVSFFPSIY-----542  
 QY 647 SPTFNNSSWMLDMSVNLGCI PKETIGSMPTFIINLGHNDISGIPDEVDLGLNLT 706  
 Db 543 -----LNNRLNGTILPEIGRLKELHMLDLRRNFTGTIPDSISGLDNLBVL 588  
 QY 707 LDISNKLDRIPQAMSLTMTLEIDSNNNLSGPIPEMGQFETPPAKLNNPGLCGYP 766  
 Db 589 LDISNKLDRIPQAMSLTMTLEIDSNNNLSGPIPEMGQFETPPAKLNNPGLCGYP 766  
 QY 767 LPRCD-----PSNADGYAHQHSRRRPSIAGSVAMGLLFSFVCFGLIVLVEREKKR 821  
 Db 649 DSPCCVLMNMLNPKSGSSRRNNNGKFRGSIVVLISLIGTILLSTLVL-----RI 702  
 QY 822 RKKEALEMVAEFGHNGSGDPTANTTMMKLTGYKEALS-INLAAPER-PLRKTFTADLLQA 879  
 Db 703 SRKQVD-----DRINVDERTISGVSKALGPSKIVLFHSCGCKDSEVELKS 750  
 QY 880 TNGFHNDSLIGSGFGDVYKAILKDSGAVAIKKLHVSOGGOREFAEMETIGKIKHRL 939  
 Db 751 TNNFSQANITIGCGFGFLVYKANPPDSKAAVRLSDCCQEMERFQAEVALSRAEHKRL 810  
 QY 940 VPLIGYKQVDEBRLVNEVWKYSLSDVLTQDPKKGCVKLKSTRKIALSARGAPLPH 999  
 Db 811 VSLQGYCKHGNDRLLYSFWMENGLDYLWHLHERVDGNTLIMVRLKINGARGALAYLHK 870  
 QY 1000 NCSPHIIRHDKSNVLTLENNLEARYSDGMAKLSAMDTHLSVSTLGTGTYVDPPEYQ 1059  
 Db 871 VCEENVVHRVYKSNLIDBEKFEHLADFLARLLRPYDTHV-TDLVGTIGYIPPEYSQ 929

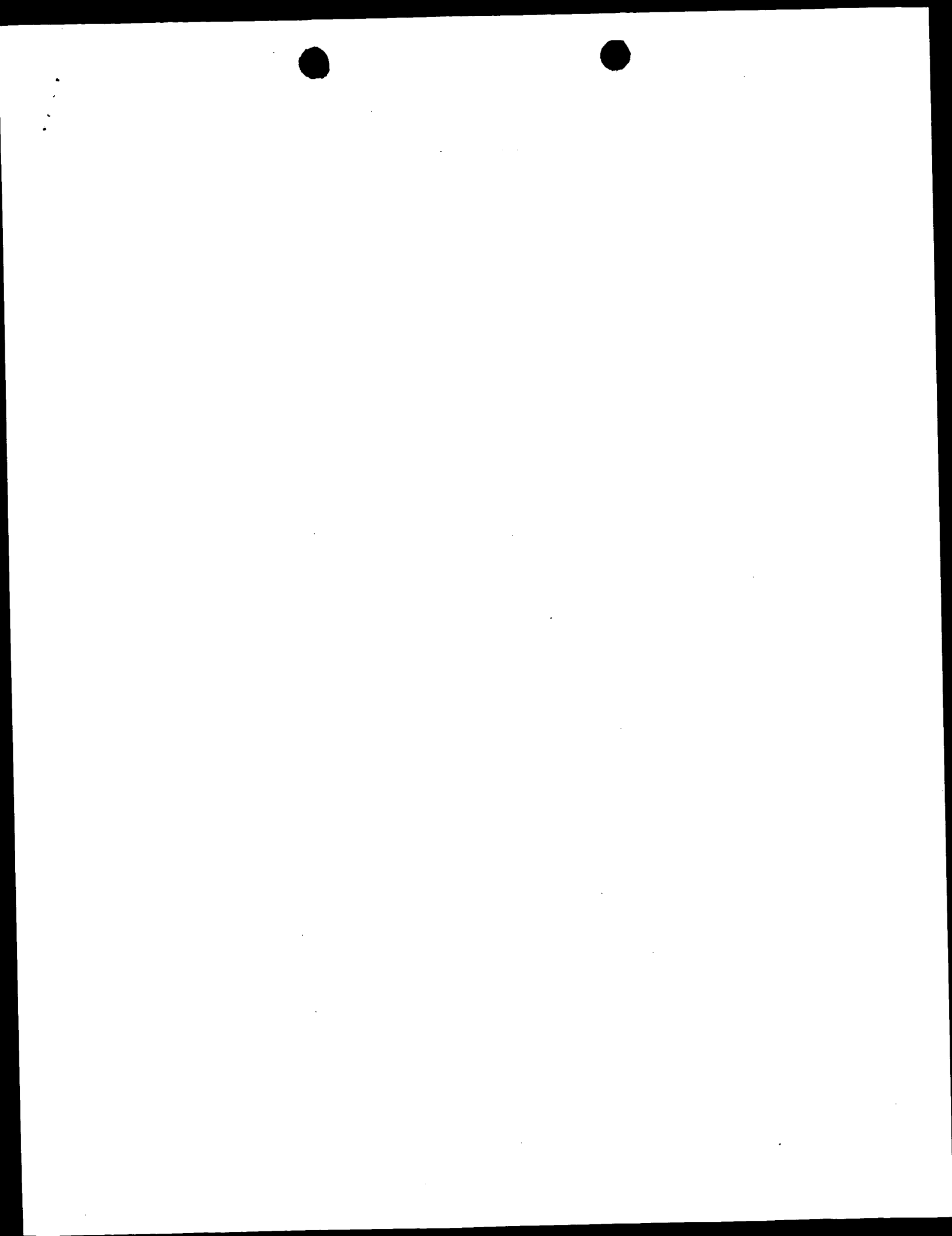
QY 1060 SFRCTKDDVSYGVVLELLTGKRPD-SPDFGDNLVGWKQ-HAKLRISVDPELM 1117  
 DB 930 SIAACRGVVSFGVVLLELVGRPAVCKCKSCRDVSRVQWAKAKREABLDITTI- 988  
 QY 1118 KEDPALEIELLOHLKVAACLDDBMRPRPTWQVAMFKEI 1158  
 DB 989 -RENNERTVLEMLIACKCIDHEPRRRLIEVVWLEDL 1028  
 RESULT 14  
 ID Q9LVPO PRELIMINARY; PRT; 1102 AA.  
 AC Q9LVPO;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)  
 DE Receptor-like protein kinase.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
 ON NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLUMBIA;  
 RX MEDLINE=20181125; PubMed=10718197;  
 RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,  
 RA Tabata S.,  
 RA Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence  
 RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC  
 RT clones.";  
 RL DNA Res. 7:31-63(2000).  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL: AB019227; BAA96896.1; -;  
 DR InterPro: IPR000719; Euk\_Pkinase.  
 DR InterPro: IPR001611; LRR.  
 DR InterPro: IPR003592; LRR\_out.  
 DR InterPro: IPR002290; Ser\_thr\_kinase.  
 DR InterPro: IPR01245; Tyr\_kinase.  
 DR Pfam: PF00560; LRR; 23.  
 DR Pfam: PF00069; Pkinase; 1.  
 DR PRINTS: PRO0019; LEURICRPT.  
 DR PRINTS: PRO0109; TYRKINASE.  
 DR ProDom: PD000001; Euk\_Pkinase; 1.  
 DR SMART: SM00370; LRR; 22.  
 DR PROSITE: PS50011; PROTEIN KINASE DOM; 1.  
 DR PROSITE: PS00108; PROTEIN KINASE ST; 1.  
 DR ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
 YR -  
 YR SEQUENCE 1102 AA; 120476 MW; 6CDP852ADD3D254D CRC64;  
 Query Match 22.3%; Score 1380.5; DB 10; Length 1102;  
 Best Local Similarity 32.0%; Pred. No. 1.7e-81;  
 Matches 387; Conservative 179; Mismatches 470; Indels 173; Gaps 31;  
 QY 1 MHTPSSFEIVTTLFFEFPSLSPQSPQSLYREIHQILSFKVLDP-KNLLPDWSSNK 59  
 DB 5 MMLAVFFSLILLILI-----SETTGALNGQYLLEIKSFVAKQKLRWMSND 55  
 QY 60 N-PCTFDGVTTC---RDDKVTSIDLSKPLNVGSVSSSLSTGLSLPFSNHNIS 114  
 DB 56 SVPCGWTGVCNSVSDPEVLSTINLSMWL-----SGKSPS 92  
 QY 115 VSGFKCSASITSLDSRNLSGVTTTLTSGSCGLKPLNVSNTLDPFGKVSGLKINS 174  
 DB 93 IGLI---VHLKQLDLNGLSGKIP--KEIGNCSSLEILKLNQPDGIPFIEIG-KLVS 146  
 QY 175 LEVLDSANSISG-----ANVGVMLSDGGELKHLA---ISGNKIS 213  
 DB 147 LEMLIYNNRISGLSVEIGNLISLQVLVYSNNISGQ-PRSIGNLKRLTSPRAGQNMIS 206  
 QY 214 GDV--DVSRVNLPLDVSSNNFSTGIP-FLGPCSALQHLIDISGNLSDGFSRATSTCTE 270  
 DB 207 GSIPSEIGCESIVMLGLAQNOLSGELPKFEGMLKKLSQVILWENEFSGFIPHEISNCTS 266

QY 271 LKLNISNOFVPIPLP--PLKSLQVLSLENKFTGEIPFLSGACDITLGLDSGNH 328  
 DB 267 LETLALYKNQVLGPPIKELGDIQSLFELYLYNGANGNITPRI-GNLSAIAIDPSENAL 325  
 QY 329 YGAVPPFPSCSLLESIALSSNNFSGELPMDTLLKMRGLKVLDSFNEFSGELPSLNL 388  
 DB 326 TGEIPELGNIELELTVFENQLTGTIPE--LSTLKNLSKLDSINALTGPILPGFOYL 384  
 QY 389 SASLLTLDSSNNFSGPIPLNLCQNKNTLQELYLQNNGFTGKIPPTLSCSELVSLHS 448  
 DB 385 -RGLFMQLQFQNSLSTGTPPKL--GWYSDLVWMDSDNHSRIPBYCLHNMNITLNG 441  
 QY 449 FNYLSGTPESLISKLKDLKMLNMEGEI-POELMYVKTLETLILDNDITLGEIPSEL 508  
 DB 442 TNNLSGNTPIGTCTKTVGLRLARRNNLVGRFNSNLCQVNVATIELGQNRFRGSIPREV 501  
 QY 509 SNETNLWISLNNRLTGEIPKWTIGLENLAILKLSNNSFSGNIPDELDCSLWLDEN 568  
 DB 502 GNCALQRLQDLADNGFGEIPREIGMLSQGLTINISSNKLTVGEIPSEIRCKLQRLDMC 561  
 QY 569 TNLFNGTIPAMFQSGKIAMFIAKRYVYIKNDGKKKECHGAGNILEFQIRSQLR 628  
 DB 562 CNFSGTLPSEV-----GSLYQL-----ELLK 583  
 QY 629 LSTRNPNITSRVYGGHTSPFDNNGSMFLDMSYNNLSGTYIREIGSMPTLFI-LNLGH 687  
 DB 584 LSNNN-----LSGTIPVALGNLSRLTELQMGNGNLFNSGIPRELSLGLQIALNLST 635  
 QY 688 NDIGSIPDEVGDLRGAILDLSNKKLDGRIPOAMSAITWTELDSSNNLSGPIPEMG 747  
 DB 636 NKLTVGEIPPELSNLMVLEFLLNNNNLSGELIPSSFANLSLGLNFSYNLTGPIPL-- 693  
 QY 748 PETPPEAKFLNPNGLCGVPLPRC-----DPSNADGYAHHQRSHGRPPASLAVAGML 801  
 DB 694 -RNISMSPFIGNELCGPLNQCIQTQPFASQSTG-----XPQGRSKIIAIT 742  
 QY 802 FSVFCIFGLIIVGREMRKRRKKEALEMYAEHGNGSDRTANTNKKLGVKALINL 861  
 DB 743 AAVIGVSLMLALIVLMRPRPVTVASSAODG-----QPSMSL 782  
 QY 862 AAFEPKLRKLTPEADLQATNGFNHDSLISGFGDGVYKALIKDSAVAIIKL--HVSQ 919  
 DB 783 DIFPPKSGEFTQDLVAATNDPDESFFVGRGACGVYAVLPAGYTLAVKKLASNHEGN 842  
 QY 920 G---DREFMAEMETICKIKNRLVPLIGCYKGDERRLLVNEVMYKGSLEDTLOPKKGV 976  
 DB 843 NNNVDSFPAELITLGNIRHRIVKLHGFCNMQSGNLLLEYMPKSGSLGELIDP---SC 899  
 QY 977 KKLSTRKIAIGSARGLAFLLHNGSPHIIHRDMKSNVLDENLEARVSDFGMARLMSA 1036  
 DB 900 NLDWSKRFKIALGAAGLAVLHDDCKPRTFHRDLSNNILLDDKREAVGDFGLAKVID- 958  
 QY 1037 MDTHLSVSTILAGPGVYPEYOSFPCSTKGDVSYGVVLELLTGKRPDSDPDGNNL 1096  
 DB 959 MPKSKMSALASGYIAEYAVYMKVTEKSDIYSGVVLELLTGKRPVQPIDG-GDV 1017  
 QY 1097 VGVWKQHAQ--LRISVPPELMKEPDALEILOHLKVAACLDDBMRPRPTWQVAM 1154  
 DB 1018 VNNVRSYTRDALSGVLDARLTLEDERIVSHMLTVLKTALCTSVSPARSMQGVLM 1077  
 QY 1155 FKEIQAGSG 1163  
 DB 1078 LIESERSEG 1086  
 RESULT 15  
 ID Q9CA77 PRELIMINARY; PRT; 1079 AA.  
 AC Q9CA77;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)

DE Putative receptor protein kinase.  
GN FLM20.4.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eubacteriales; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsidae.  
NCBI\_TaxID=3702;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CY. COLUMBIA;  
RX MEDLINE=21016719; PubMed=11130712;  
RA Theologis A., Becker J.R., Palm C.J., Federpiet N.A., Kaul S.,  
RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.V.,  
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
RA Dunn P., Eguen P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
RA Gill J.B., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,  
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,  
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
RA Lin X., Liu S.X., Liu Z.A., Lucos J.S., Maiti R., Marziani A.,  
RA Miltcher J., Miranda M., Nguyen M., Nieman W.C., Osborne B.I.,  
RA Pail G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
RA Sakano H., Salzer S.L., Schwartz J.R., Shum P., Southwick A.M.,  
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
RA Utechtback T., Van Aken S., Vaynsberg M., Vysotskaia V.S., Walker M.,  
RA Wu D., Yu G., Fraser C.M., Venter U.C., Davis R.W.,  
RA "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
RT thaliana".  
RL Nature 408:816-820(2000).  
DR EMBL, AC011765; AAC52362.1;  
DR InterPro: IPR000719; Euk\_pkinase.  
DR InterPro: IPR001611; LRR.  
DR InterPro: IPR003591; LRR\_Typ.  
DR InterPro: IPR002290; Ser\_Thr\_pkinase.  
DR Pfam: PF001245; Tyr\_pkinase.  
DR Pfam: PF00560; LRR\_17.  
DR Pfam: PF00069; Kinase; 1.  
DR PRINTS: PR00019; LEURICHRPT.  
DR PRINTS: PR00109; TYRKINASE.  
DR ProDom: PD000001; Euk\_pkinase; 1.  
DR SMART: SM00369; LRR\_Typ; 7.  
DR SMART: SM00220; S\_TKc; 1.  
DR SMART: SM00219; TYKc; 1.  
DR PROSITE: PS00107; PROTEIN KINASE ATP; UNKNOWN\_1.  
DR PROSITE: PS00101; PROTEIN KINASE DOM; 1.  
DR PROSITE: PS00108; PROTEIN KINASE ST; 1.  
DR Kyr ATP-binding; Kinase; Recptor; Transferase.  
DR Kyr SEQUENCE 1079 AA; 118675 MW; A19918DD9B8D4 CRC64;  
Query Match 22.3%; Score 1380; DB 10; Length 1079;  
Best Local Similarity 32.2%; Pred. No. 1.8e-81;  
Matches 392; Conservative 204; Mismatches 406; Indels 214; Gaps 44;  
QY 26 ASFSGLYEIRHQLISFKVLPDKN-----LLPDWSKNKP--CTPDGVC--RDDKVTYS 76  
DB 5 AVAGDLDDBREVLTLSKYSLRKNPQNGLYTEKMNQDVQWCPPIICTPGQSRVYG 64  
QY 77 IDS-----SKPLNVSFAVSSSLITGLSESLFSLNSHINSVSGFKCSASLTLDLSRN 132  
DB 65 INLTSTISGPFKPKFSALT-----ELTLDLSRN 94  
QY 133 SLSGPTTLTSLGSCSGKFLNVSNTLD-----FPKVSGLKLSLEVLDSANSTISGA 188  
DB 95 TIEGRP--DLSRCHNLKHLNLSNILEGELSLPG-----LSNLEVLDSLNRITGD 145  
QY 189 NVGVWVLSGCGELKHLAISGNKISGVDV--VSRVNLFLPVDVSNFNS-----TGI----- 238  
DB 146 IGSFPLF--CNSLVVAMLSSTNFTGRIDTDFNGCRNLKTYVDSNPSFSGVWTFGRGLV 203  
QY 239 -----PLGDCSALQHLDISGNKLSGDFSRALSTCTELKILNINNOFV 282  
DB 204 ERSVADNHLGNISASWFRNGCT--LQMLDLSGNAFGGEFPQGVSNCONLNVLMKNKFT 262

QY 283 GPIPL--PLKSHOYLSLAENKPTGEIPDLGSCADTLTGDLGNNHYGAVPFFGSCS 340  
DB 263 GNIPARIGISLSLKGILGNNTFSRIDPPTLNLTN--LVFIDLSENKRGQIOEIFGRFT 321  
QY 341 LLESLLASNNPSGELPMDTLLKRGKLVLSFNEFSGELPESLTNLSASLTLDLSN 400  
DB 322 QVKYLVLANSVYGGINSNLIKPNLSRIDLGYNPFSGQLPTEISQI----- 369  
QY 401 NFGSILPMLQNPNTLOELYLQNNGTGKIPTPLNSCSELVSLHLSFNYSLSGTPSSL 460  
DB 370 -----QSLKFLTLAVNNSGDIPOEYGMPELQALDLSFNKLTGTSIASF 414  
QY 461 GSLSIKRIDKLMNLNBEIPQELMYVTLTETLLIDFNDLGEIPSGLSNCTNLTSLSS 520  
DB 415 GKLTSI--LW-----LNLANSLSGELPREIGCTSLMFWVA 450  
QY 521 NNRLTGEIPKWTIGRL--EMLALILKSNNSFSGNIPDELGDRSL--IWLDTNLTNFGTIPA 578  
DB 451 NQSLGRFHPPELTRGNSPSPTEFVNQNKDILTAGSGECLAMKRWIPARFPFN--FVYA 509  
QY 579 AMFKSGKIAAFIAGKRYVYIKDGMKKECHGAGNLEFQGRSEQLRLSTRPCNT 638  
DB 510 ILTKSGSLMDHY-----LKGVLFPVC-SAGST-----VRLKISAY-----LOLS 551  
QY 639 SRVYGHTSPFTDNNSGMPLDMSYNNLSGYIPKEIGSMPLYFLMIGHNDISGSPDEY 698  
DB 552 GNKFSGETIPASISQMDRLSTLHGFNEFEKLPPEIGQLPLAF--LNLTNNSGELPQEI 610  
QY 699 GDLRGLNILDSSNKLDRIPQMSALITMLTDLNNSN--LSGPIPEWQFTFPAPKEL 757  
DB 611 GNLKCLQNLDSLSPNPSGAPPTSLNDLNLSEKSNISYNPISGALPTTQVATFPKDSPL 670  
QY 758 NNPLCGEPLPRCDPSNADGYAHRQSH--GRRPASLA--GSYAMGLPFGVCIPLDI 811  
DB 671 GNP-LIRP-----SFFQSGNNTKRIKISNOVLGNRPRTLLIISLALAF-----LACIV 721  
QY 812 LVGERMKRRKKELEEMVABEGHNSGRTAN--NTNWKLTGYKALSLNLAPEKPL 868  
DB 722 VSGIVLVWVYASREAITD--LDSKTRHDMTSSGSSSPW--LSGKIKYIRLDKSTF----- 775  
QY 869 RKLTPADLQATNGFNHDSLISGCGFVYKAILKDGSAVAILKLIHVSQGDDEFAEM 928  
DB 776 ---TYADILKATSNFSEKRVVGRGVYRGVLPDGEKVAVKLQREGTAERKEFAEM 832  
QY 929 ETI-----GKIKRNLVPLIGCTKVGDERLLVNEVWKXSLDEVLODPKKGVLKLTSTR 983  
DB 833 EYLSANAFEDMAHPNLVRLYGMCLDSEKILVHEYMGSGSLEELITD-----KITLQWK 886  
QY 984 RKIATGS--ARGLAFLHNHNSPHIIRHMKSSNVLDENLFAVSDTGAMRLNSAMDTHL 1041  
DB 887 KRDIATDVARGLVFLHHCYPSIYHRVYKASNVLLDGHGAAVATDGLARLLNVGDSHV 946  
QY 1042 SVSTLAGTGYTPPEYVYVFCSTKGDVYSYGVLLBELTGKRPDPSPDGDNLVGWVK 1101  
DB 947 S-TVAGTGYVAPPYGQVQATTRGDVYSYGVLLTRELAGRAVDG--GECILVEMAR 1002  
QY 1102 QHAKRLISDVEPELM--KEDPALTEILQHKVANAACLDPRAMRRPTVQVAMFKEI 1158  
DB 1003 RYMTGNMTRAKSPITLSGTRKGNAB--QMTLELKIGVCTADHPARPNMKEVTLAMLVKI 1061  
QY 1159 ---OAGGINDSGST 1170  
DB 1062 SGKALFNLGLSSGGYI 1077

Search completed: March 10, 2003, 18:18:41  
Job time : 63.1415 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 18:16:35 ; Search time 19.0143 Seconds

(without alignments)  
2608.861 Million cell updates/sec

Title: US-09-823-394-2

Sequence: 1 MKTFSSFFSVTLTFFSF.....GFSTIEMVDMSTKEVPEGKL 1196

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1311	21.2	999	RLKS_ARATH	P47735 arabidopsis
2	1225.5	19.8	980	Q94YQ8 ARATH	Q94YQ8 arabidopsis
3	834	13.5	942	TMKL_ARATH	P43298 arabidopsis
4	639	10.3	674	TMKL_ARATH	P33543 arabidopsis
5	543.5	8.8	389	NAK_ARATH	P43293 arabidopsis
6	528	8.5	901	CR14_MAIZE	O24585 zea mays (m
7	508	8.2	410	APRA_ARATH	Q06548 arabidopsis
8	507.5	8.2	412	APRA_ARATH	P46573 arabidopsis
9	507	8.2	817	KPRO_MAIZE	P17801 zea mays (m
10	498	8.1	849	SRK6_BRAOL	Q09092 brassica ol
11	401.5	6.5	372	D100_ARATH	Q00874 arabidopsis
12	393.5	6.4	582	SHO2_HUMAN	Q94913 homo sapien
13	392	6.3	710	IRAI_MOUSE	Q62406 mus musculu
14	384.5	6.2	501	KEPL_MOUSE	Q06552 drosophila
15	378.5	6.1	582	SHO2_MOUSE	Q08520 mus musculu
16	378	6.1	496	TMW_ARATH	Q94913 arabidopsis
17	377.5	6.1	712	IRAI_HUMAN	P5617 homo sapien
18	341.5	5.5	821	CR1_ARATH	Q06509 arabidopsis
19	338.5	5.4	1584	KTKI_DICDI	P24466 ascaris
20	331.5	5.4	1839	CHAO_SACAL	P12024 drosophila
21	329.5	5.3	1315	CHAO_DROME	Q01631 neurospora
22	328	5.3	2300	CHAO_NEUCR	P33334 phaseolus v
23	327	5.3	342	PG13_PHAVU	P58823 phaseolus v
24	324	5.2	342	PG13_PHAVU	Q01513 podopora a
25	323	5.2	342	PG12_PHAVU	Q01513 podopora a
26	320.5	5.2	2145	PG12_PHAVU	Q01513 podopora a
27	303.5	4.9	330	PG11_ARATH	Q06513 pyrus commu
28	300	4.9	330	PG12_ARATH	Q06513 pyrus commu
29	297.5	4.8	330	PG12_ARATH	Q06513 pyrus commu
30	296	4.8	2493	BRAT7_CHICK	P49606 utillago ma
31	292	4.7	1692	BRAT7_CHICK	O42422 gallus gall
32	290.5	4.7	1692	BRAT7_CHICK	P49606 utillago ma
33	290	4.7	825	TRKC_PIG	P24786 sus scrofa

34	289	4.7	1002	1	EPB5_CHICK	O07497 gallus gall
35	288	4.7	977	1	EPB2_MOUSE	O03145 mus musculu
36	287.5	4.6	984	1	EPB1_HUMAN	P54762 homo sapien
37	287.5	4.6	985	1	EPB4_XENLA	Q91571 xenopus lae
38	287	4.6	902	1	EPB3_XENLA	Q91736 xenopus lae
39	286.5	4.6	976	1	EPB1_HUMAN	P21709 homo sapien
40	286.5	4.6	984	1	EPB1_RAT	P09759 rattus norv
41	286	4.6	832	1	ANR3_HUMAN	P57078 homo sapien
42	285.5	4.6	821	1	TRKB_MOUSE	P15209 mus musculu
43	285.5	4.6	821	1	TRKB_RAT	O63604 rattus norv
44	284.5	4.6	794	1	TRK1_LYMS	O76997 lymnaea sta
45	284.5	4.6	1024	1	POPC_RALSO	Q97822 ralsomonia s

## ALIGNMENTS

RESULT 1

ID RLKS\_ARATH STANDARD; PRT; 999 AA.

AC P47735;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Receptor-like protein kinase 5 precursor (EC 2.7.1.-).

GN RLKS OR A74G28490 OR F2109.180.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OK NCBI\_TaxID=3702;

OK [1]

RN SEQUENCE FROM N.A.

RP STRAIN=cv. Columbia;

RC MEDLINE=20083488; PubMed=10617198;

RX Walker J.C.;

RT "Receptor-like protein kinase genes of Arabidopsis thaliana.";

RL Plant J. 3:451-456(1993).

RU [2]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=20083488; PubMed=10617198;

RA Mayer K.F.X., Schaefer C., Wambert R., Murphy G., Voelckert G.,

RA Poll T., Duesterhoeft A., Stiekema W., Eutian K.-D., Terry N.,

RA Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M.,

RA Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M.,

RA Kreis M., Delany M., Purgdomenech P., Watson M., Schmidheini T.,

RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,

RA Vos P., Hohnselt J., Zimmermann W., Wedler H., Ridley P.,

RA Langham S.-A., McCullagh B., Bilham L., Robben J.,

RA Van der Schuren J., Grymoprez B., Chuang Y.-J., Vandenbusche F.,

RA Braeken M., Wellens I., Voet M., Bastiaens I., Aert R., Defoor E.,

RA Weitzenecker T., Botte G., Rameberger U., Hilbert H., Braun W.,

RA Holzner E., Brandt A., Peters S., Van Staveren M., Dirxse W.,

RA Moellman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,

RA Bernsmeier S., Hempel S., Feldpausch M., Lambrecht S., Van den Daele H.,

RA De Keyser A., Buysaert C., Gielen J., Villalobos R., De Clercq R.,

RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,

RA Clark L., Doggett J., Hall S., Kay M., Leonard N., McElroy K., Mayes R.,

RA Pettitt A., Rajandream M.A., Lyne M., Benes V., Reckmann T.-H.,

RA Borkova D., Bloecher H., Scharte M., Grimm M., Loehner T.-H.,

RA Dose S., de Haan M., Maarse A.C., Schaefer K., Mueller-Auer S.,

RA Gabel C., Fuchs M., Fartmann B., Grandjean K., Danner D., Herzl A.,

RA Neumann S., Argitrou A., Vitale D., Liguori R., Piravandi E.,

RA Massenet O., Outgley F., Clabaud G., Mendenhall A., Falber R.,

RA Schnabl S., Hiller R., Schmidt W., Leharz A., Aubourg S.,

RA Chefor F., Cooke R., Berger C., Montfort A., Casacuberta E.,

RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,

RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,

RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,

RA Frisman P., Haase D., Lemcke K., Mewes H.-W., Stocker S.,

RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,

RA Parnell L., Dedha N., Gnoj L., Schutz K., Huang E., Spiegel L.,

RA Sektion M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,  
RA Stenking T., Kallick J., Graves T., Harmon G., Edwards J.,  
RA Latrell P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,  
RA Nelson J., Spiehl J., Ryan E., Andrews S., Giesel C., Layman D.,  
RA Du H., Ali J., Bergdorf A., Jones K., Dione K., Cotton M., Joshi C.,  
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,  
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,  
RA Grant S., Shohdy N., Haegawa A., Rodriguez M., Hoffman J., Tili S.,  
RA Chen E., Marra M., Martienssen R., McCombie W.R.;  
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis  
thaliana.";  
RL Nature 402:769-777(1999).  
RN [3]  
RP CHARACTERIZATION.  
RX MEDLINE=9436830; PubMed=8086440;  
RA Horn M.A., Walker J.C.;  
RA "Biochemical properties of the autophosphorylation of RLK5, a  
receptor-like protein kinase from Arabidopsis thaliana.";  
CC Biochim. Biophys. Acta 1208:65-74(1994).  
CC -1- COPACOR: HAVE SIGNIFICANTLY GREATER ACTIVITY IN THE PRESENCE OF  
NM2+ THAN MG2+.  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ROOTS AND ROSETTES.  
CC -1- PERM: AUTOPHOSPHORYLATED ON SERINE AND THREONINE RESIDUES.  
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE SER/THR  
CC FAMILY OF PROTEIN KINASES.  
CC -1- SIMILARITY: CONTAINS 18 LEUCINE-RICH REPEATS (LRR).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
at the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M84660; AA32859.1; -;  
DR EMBL; AL021749; CAAL6889.1; -;  
DR EMBL; AL161572; CAB79651.1; -;  
DR InterPro: IPR000719; Euk\_pkinase.  
DR InterPro: IPR001611; LRR.  
DR InterPro: IPR003592; LRR\_out.  
DR InterPro: IPR002290; Ser\_thr\_pkinase.  
DR Pfam: PF00069; Pkinase; 1.  
DR Pfam: PF00560; LRR; 20.  
DR PRINTS: PR00019; LEURICHRPT.  
DR ProDom: PD000001; Euk\_pkinase; 1.  
DR SMART; SMO0370; LRR; 17.  
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
DR PROSITE; PS00108; PROTEIN KINASE ST; 1.  
DR TRANSFAC; PS50011; PROTEIN KINASE DOM; 1.  
KW Transferrase; Serine/threonine-protein kinase; ATP-binding;  
KW Transmembrane; Glycoprotein; Phosphorylation; Leucine-rich repeat;  
KW Repeat; Signal.  
FT SIGNAL 1 14  
FT DOMAIN 15 999  
FT TRANSMEM 622 641  
FT DOMAIN 642 999  
FT REPEAT 88 112  
FT REPEAT 114 138  
FT REPEAT 139 161  
FT REPEAT 163 186  
FT REPEAT 188 211  
FT REPEAT 235 261  
FT REPEAT 263 283  
FT REPEAT 284 306  
FT REPEAT 307 330  
FT REPEAT 332 354  
FT REPEAT 354 402  
FT REPEAT 402 426  
FT POTENTIAL 14  
FT RECEPTOR-LIKE PROTEIN KINASE 5.  
FT EXTRACELLULAR (POTENTIAL).  
FT POTENTIAL.  
FT CYTOPLASMIC (POTENTIAL).  
FT LRR 1.  
FT LRR 2.  
FT LRR 3.  
FT LRR 4.  
FT LRR 5.  
FT LRR 6.  
FT LRR 7.  
FT LRR 8.  
FT LRR 9.  
FT LRR 10.  
FT LRR 11.  
FT LRR 12.

FT REPEAT 427 450 LRR 13.  
FT REPEAT 452 474 LRR 14.  
FT REPEAT 498 522 LRR 15.  
FT REPEAT 524 547 LRR 16.  
FT REPEAT 549 567 LRR 17.  
FT REPEAT 568 592 LRR 18.  
FT DOMAIN 683 968  
FT CARBOHYD 98 98  
FT CARBOHYD 102 102  
FT CARBOHYD 150 150  
FT CARBOHYD 185 185  
FT CARBOHYD 210 210  
FT CARBOHYD 269 269  
FT CARBOHYD 282 282  
FT CARBOHYD 452 452  
FT CARBOHYD 576 576  
FT NP BIND 689 697  
FT BINDING 711 711  
FT ACT SITE 819 819  
FT MUTAGEN 711 711  
SQ SEQUENCE 999 AA; 109095 MW; F5793D89E9A0C6A7 CRC64;  
Query Match 21.2%; Score 1311; DB 1; Length 999;  
Best Local Similarity 32.4%; Pred. No. 2.5e-68;  
Matches 386; Conservative 161; Mismatches 392; Indels 254; Gaps 30;  
22 LSFQSPQSLVREIHOLISFQVDPD-KNLLPWSNKN--PCTPDGVC-RDDKYTSI 77  
11 LSTTLPISLWQDITIRQAKLSLSDPSQSLSDNDVTPCKWLGVCSDATSNVVS 70  
78 DLSKPLNVGFSAVSSLSLTGLESFLSNHNGSVG--FKCSASLTSLDLSRSL 135  
71 DLSFPLVGFPSI---LCHLPSLHSLNYSNGSLSDADFTCHNLISLDSNTLV 127  
136 GCVTTITSLGSCGLKFLNVSNTLDPFGKXSGGKANSLEVLDSANSIGANVGVWL 195  
128 GSIP-----KSLPFLP----- 139  
196 SPOGCELKHLAISGNKISGVDVSRVNLFLDVSSNNFSGTIP-FLDCSALQHLDISG 254  
140 -----NKLFLISGNLSLDTIPSPFQFRLSLINLAG 172  
255 NKLSDPSRAISTCTELKILNINSNQF-VGPIPLP--PLKSIQYSLAKNFTGSIPLFL 311  
173 NPLSGTIPASLGNVTLTKELKLAIVLFSPOQIPSOQIGTLEQVIMLGCNLVGIPLSL 232  
312 SGACDTLTGLDSGNHFGVAVPFPFGSCSLSLSSNPFSGEIPMDTLMRGLKYLD 371  
233 S-RLTSLVNLDTLFPNLTGSIPSWITQLKTVQIQLFNNSFSGELP-ESMGNTTLKFD 290  
372 LSFNPSGELPESTLNLASLLTLDLSSNPFSGPILPNLCQPKNTLOELVYQNGFTGK 431  
291 ASNKKLTGKIPDLNL--NLSSINLFEMWLEGPPESTRS--TLTSLTKLPNNRLTGV 346  
432 IPTLNSCELVLSHFNVTLSGTISLSLSLSKLRDLKLMNMLEGELPQGLMYVTKLE 491  
347 LPQLGANSPLQYVDLSYNRFSGEIPANVCGEKLKYLILINDSFGESLNNLGCKSLT 406  
492 TLILDFNDLGEIPSGISNCTNLNMTLSNNRLTGEIPMIGRELENTALIKLNNSPSGN 551  
407 RVRSNKSLSGQIPHFGWGLPRLSLELSDNSTGSIPTTIIGAKLSLRLSKRFSGS 466  
552 IPDELGDCSLIWLIDNTLNFNGTIPAMFKOSGLIANFLAGKRYVYIKNDGKKECHG 611  
467 IPNIGSLNGIIRISGAKNDFSGEIPESL----- 495  
612 AGNLLFQGRSQNLRLSTRNCITTSRYGHTSPPTDNNGSMFLLMSVYMLSGYIP 671  
496 -----VTKQLSR-----LDLSKNQSLSGEIP 516  
672 KEIGSMPLYFLINLGHNDISGIPDEYDRLGNIILDSNKLDRIPAMSNLTMVTEI 731  
517 RELGKQNLNLEINLANNHLSGEIPKEVGILLPVANIYDLSNQSGLIPLQLNL-KLVNL 575



QY 732 DLSNNNLSGPIPEKGEETPEPPAFELNPNGLCGYPLRCDPSNADGYAHQSHGRBPAS 791  
 Db 576 NLSYTHLSGKIPLYANXVY-AHFIFGNPGIC-----VLDLGLCR-----K 615  
 QY 792 LAGSVAMG-----LFSFVCIPLGLLVGRMRGRKKKAEAEEMTAEBHGSGDRT 842  
 Db 616 ITRKKNIGYVILITLITFLMGLVFPVGIWFIKCRRLALKSTTLA----- 663  
 QY 843 ANNTNMKLTGVKEALSINLAFFERPLKLTFF-----ADLLQATNGFHNLSLIGSGGFYV 897  
 Db 664 ---SKM-----RSFHKLHFSHEHIALDLDEKN-----VIGFGSGGY 697  
 QY 898 YKALLKXGSAVAIKLTHVSGQDRE-----FMAEMETIGIKIKARNVPLLYGCK 947  
 Db 698 YKVALRGEVVAIVKRLKSKYKGVGDDEYSDSLNRDVFALAEVEITLIRKHSIYRLMCCS 757  
 QY 948 VGERLLLVNEMKYSLEDTLOPKKGVYKLKSTRKXIAIGSARLALHNCSPHIIH 1007  
 Db 758 SGDCKLIVYEMRNGSLADVLHGRKSGVYLGWPERLRIALDAEGLSYLHHDVCPPIVH 817  
 QY 1008 RDMKSSVNLIDENLEAFVSDFGMARL--MSAMDTLHLSVSTLAGTPGVPEYOSFRCST 1065  
 Db 818 RDVSSNILLDSYGAIVADFGIAKYGQSGSKTPEAMSGINGSCYIAPVYTLRYNE 877  
 QY 1066 KGVYSGVVLLELTGKRTSPDGDNNLVGVV-KQIAKLRISDVPELMKEDPALE 1124  
 Db 878 KSDIITSGVVLLELTGKRTSPDGDNNLVGVV-KQIAKLRISDVPELMKEDPALE 1124  
 QY 1125 IELLQHLKVAIVACLDRAWRRPTMVOMAFKRIQAGSGIDSGITRISLEDOG 1177  
 Db 934 EISKVIYHGLTSPPLNRPBMRKVIVLMQVSGAVPCSPNTSKSKTKG 986

## RESULT 2

CLVL ARATH STANDARD; PRT; 980 AA.  
 ID CLVL ARATH STANDARD; PRT; 980 AA.  
 AC Q9SY08; 004380; 09LOT2;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein kinase CLAVATA1 precursor (EC 2.7.1.1-).  
 GN CLVL OR AT1G75820 OR T4012.5.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyle; Rosidae;  
 OC eurosida II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_Taxid=3702;  
 KY 111  
 KX SEQUENCE FROM N.A.  
 RC STRAIN=cv. Landsberg erecta;  
 RC MEDLINE=97304386; PubMed=9160749;  
 RA Clark S.E., Williams R.W., Meyerowitz E.M.;  
 RT "The CLAVATA1 gene encodes a putative receptor kinase that controls  
 RL shoot and floral meristem size in Arabidopsis.",  
 RL Cell 89:575-585 (1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE=99178804; PubMed=10080719;  
 RA Williams R.W., Clark S.E., Meyerowitz E.M.;  
 RT "Genetic and physical characterization of a region of Arabidopsis  
 RL chromosome 1 containing the CLAVATA1 gene.",  
 RL Plant Mol. Biol. 39:171-176 (1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=cv. Columbia;  
 RC MEDLINE=21016719; PubMed=11130712;  
 RA Theologis A., Becker U.R., Palm C.J., Federspiel N.A., Kaul S.,  
 RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,  
 RA Buehler E., Chan A., Chen Q., Chen H., Cheuk R.F., Chin C.W.,  
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.R., Dewar K.,  
 RA Dunn P., Etgu P., Feldlyum T.V., Feng J.D., Fong B., Fujii C.Y.,  
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,

RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
 RA Kim C.J., Koo H.L., Kremetskaia I., Kutz D.B., Kwan A., Lam B.,  
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maltz R., Marziani A.,  
 RA Miltsecher J., Miranda M., Nguyen M., Nieman W.C., Osborne B.T.,  
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shin P., Southwick A.M.,  
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
 RA Uteback T., Van Aken S., Vayenberg M., Vysotskaia V.S., Walker M.,  
 RA Wu D., Yu G., Fraser C.M., Venter R.W.;  
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
 RL thaliana.";  
 RL Nature 408:816-820 (2000).  
 CC - FUNCTION: ACTS WITH CLV3 TO CONTROL THE BALANCE BETWEEN MERISTEM  
 CC CELL PROLIFERATION AND DIFFERENTIATION. MAY ACT WITH CLV3 AS A  
 CC LIGAND-RECEPTOR PAIR IN A SIGNAL TRANSDUCTION PATHWAY.  
 CC COORDINATING GROWTH BETWEEN ADJACENT MERISTEMATIC REGIONS.  
 CC - SUBUNIT: MULTIMER (POTENTIAL).  
 CC - TISSUE SPECIFICITY: IN A CENTRAL REGION OF THE SHOOT AND IN EARLY  
 CC FLOWER MERISTEMS.  
 CC - SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC - SIMILARITY: CONTAINS 18 LEUCINE-RICH REPEATS (LRR).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; U96879; AA58929.1; -  
 DR EMBL; AF049870; AA02501.1; -  
 DR EMBL; AC007396; AA26772.1; ALT INIT.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR003592; LRR\_out.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF00560; LRR; 18.  
 DR PRINTS; PR00019; LRRICHRPT.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00370; LRR; 18  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Receptor; Differentiation; Signal; Repeat; Leucine-rich repeat;  
 KW Transferase; Serine/threonine-protein kinase; ATP-binding;  
 KW Phosphorylation; Transmembrane.  
 FT SIGNAL 1 24  
 FT CHAIN 25 980  
 FT DOMAIN 25 638  
 FT TRANSMEM 639 659  
 FT DOMAIN 660 980  
 FT REPEAT 93 119  
 FT REPEAT 143 193  
 FT REPEAT 168 193  
 FT REPEAT 240 264  
 FT REPEAT 265 288  
 FT REPEAT 289 312  
 FT REPEAT 314 336  
 FT REPEAT 337 360  
 FT REPEAT 361 384  
 FT REPEAT 386 408  
 FT REPEAT 409 432  
 FT REPEAT 455 479  
 FT REPEAT 480 503  
 FT REPEAT 505 527  
 FT REPEAT 528 551  
 FT REPEAT 552 575  
 FT LRR 1.  
 FT LRR 2.  
 FT LRR 3.  
 FT LRR 4.  
 FT LRR 5.  
 FT LRR 6.  
 FT LRR 7.  
 FT LRR 8.  
 FT LRR 9.  
 FT LRR 10.  
 FT LRR 11.  
 FT LRR 12.  
 FT LRR 13.  
 FT LRR 14.  
 FT LRR 15.  
 FT LRR 16.  
 FT LRR 17.  
 FT LRR 18.  
 FT LRR 19.  
 FT LRR 20.  
 FT LRR 21.  
 FT LRR 22.  
 FT LRR 23.  
 FT LRR 24.  
 FT LRR 25.  
 FT LRR 26.  
 FT LRR 27.  
 FT LRR 28.  
 FT LRR 29.  
 FT LRR 30.  
 FT LRR 31.  
 FT LRR 32.  
 FT LRR 33.  
 FT LRR 34.  
 FT LRR 35.  
 FT LRR 36.  
 FT LRR 37.  
 FT LRR 38.  
 FT LRR 39.  
 FT LRR 40.  
 FT LRR 41.  
 FT LRR 42.  
 FT LRR 43.  
 FT LRR 44.  
 FT LRR 45.  
 FT LRR 46.  
 FT LRR 47.  
 FT LRR 48.  
 FT LRR 49.  
 FT LRR 50.  
 FT LRR 51.  
 FT LRR 52.  
 FT LRR 53.  
 FT LRR 54.  
 FT LRR 55.  
 FT LRR 56.  
 FT LRR 57.  
 FT LRR 58.  
 FT LRR 59.  
 FT LRR 60.  
 FT LRR 61.  
 FT LRR 62.  
 FT LRR 63.  
 FT LRR 64.  
 FT LRR 65.  
 FT LRR 66.  
 FT LRR 67.  
 FT LRR 68.  
 FT LRR 69.  
 FT LRR 70.  
 FT LRR 71.  
 FT LRR 72.  
 FT LRR 73.  
 FT LRR 74.  
 FT LRR 75.  
 FT LRR 76.  
 FT LRR 77.  
 FT LRR 78.  
 FT LRR 79.  
 FT LRR 80.  
 FT LRR 81.  
 FT LRR 82.  
 FT LRR 83.  
 FT LRR 84.  
 FT LRR 85.  
 FT LRR 86.  
 FT LRR 87.  
 FT LRR 88.  
 FT LRR 89.  
 FT LRR 90.  
 FT LRR 91.  
 FT LRR 92.  
 FT LRR 93.  
 FT LRR 94.  
 FT LRR 95.  
 FT LRR 96.  
 FT LRR 97.  
 FT LRR 98.  
 FT LRR 99.  
 FT LRR 100.  
 FT LRR 101.  
 FT LRR 102.  
 FT LRR 103.  
 FT LRR 104.  
 FT LRR 105.  
 FT LRR 106.  
 FT LRR 107.  
 FT LRR 108.  
 FT LRR 109.  
 FT LRR 110.  
 FT LRR 111.  
 FT LRR 112.  
 FT LRR 113.  
 FT LRR 114.  
 FT LRR 115.  
 FT LRR 116.  
 FT LRR 117.  
 FT LRR 118.  
 FT LRR 119.  
 FT LRR 120.  
 FT LRR 121.  
 FT LRR 122.  
 FT LRR 123.  
 FT LRR 124.  
 FT LRR 125.  
 FT LRR 126.  
 FT LRR 127.  
 FT LRR 128.  
 FT LRR 129.  
 FT LRR 130.  
 FT LRR 131.  
 FT LRR 132.  
 FT LRR 133.  
 FT LRR 134.  
 FT LRR 135.  
 FT LRR 136.  
 FT LRR 137.  
 FT LRR 138.  
 FT LRR 139.  
 FT LRR 140.  
 FT LRR 141.  
 FT LRR 142.  
 FT LRR 143.  
 FT LRR 144.  
 FT LRR 145.  
 FT LRR 146.  
 FT LRR 147.  
 FT LRR 148.  
 FT LRR 149.  
 FT LRR 150.  
 FT LRR 151.  
 FT LRR 152.  
 FT LRR 153.  
 FT LRR 154.  
 FT LRR 155.  
 FT LRR 156.  
 FT LRR 157.  
 FT LRR 158.  
 FT LRR 159.  
 FT LRR 160.  
 FT LRR 161.  
 FT LRR 162.  
 FT LRR 163.  
 FT LRR 164.  
 FT LRR 165.  
 FT LRR 166.  
 FT LRR 167.  
 FT LRR 168.  
 FT LRR 169.  
 FT LRR 170.  
 FT LRR 171.  
 FT LRR 172.  
 FT LRR 173.  
 FT LRR 174.  
 FT LRR 175.  
 FT LRR 176.  
 FT LRR 177.  
 FT LRR 178.  
 FT LRR 179.  
 FT LRR 180.  
 FT LRR 181.  
 FT LRR 182.  
 FT LRR 183.  
 FT LRR 184.  
 FT LRR 185.  
 FT LRR 186.  
 FT LRR 187.  
 FT LRR 188.  
 FT LRR 189.  
 FT LRR 190.  
 FT LRR 191.  
 FT LRR 192.  
 FT LRR 193.  
 FT LRR 194.  
 FT LRR 195.  
 FT LRR 196.  
 FT LRR 197.  
 FT LRR 198.  
 FT LRR 199.  
 FT LRR 200.  
 FT LRR 201.  
 FT LRR 202.  
 FT LRR 203.  
 FT LRR 204.  
 FT LRR 205.  
 FT LRR 206.  
 FT LRR 207.  
 FT LRR 208.  
 FT LRR 209.  
 FT LRR 210.  
 FT LRR 211.  
 FT LRR 212.  
 FT LRR 213.  
 FT LRR 214.  
 FT LRR 215.  
 FT LRR 216.  
 FT LRR 217.  
 FT LRR 218.  
 FT LRR 219.  
 FT LRR 220.  
 FT LRR 221.  
 FT LRR 222.  
 FT LRR 223.  
 FT LRR 224.  
 FT LRR 225.  
 FT LRR 226.  
 FT LRR 227.  
 FT LRR 228.  
 FT LRR 229.  
 FT LRR 230.  
 FT LRR 231.  
 FT LRR 232.  
 FT LRR 233.  
 FT LRR 234.  
 FT LRR 235.  
 FT LRR 236.  
 FT LRR 237.  
 FT LRR 238.  
 FT LRR 239.  
 FT LRR 240.  
 FT LRR 241.  
 FT LRR 242.  
 FT LRR 243.  
 FT LRR 244.  
 FT LRR 245.  
 FT LRR 246.  
 FT LRR 247.  
 FT LRR 248.  
 FT LRR 249.  
 FT LRR 250.  
 FT LRR 251.  
 FT LRR 252.  
 FT LRR 253.  
 FT LRR 254.  
 FT LRR 255.  
 FT LRR 256.  
 FT LRR 257.  
 FT LRR 258.  
 FT LRR 259.  
 FT LRR 260.  
 FT LRR 261.  
 FT LRR 262.  
 FT LRR 263.  
 FT LRR 264.  
 FT LRR 265.  
 FT LRR 266.  
 FT LRR 267.  
 FT LRR 268.  
 FT LRR 269.  
 FT LRR 270.  
 FT LRR 271.  
 FT LRR 272.  
 FT LRR 273.  
 FT LRR 274.  
 FT LRR 275.  
 FT LRR 276.  
 FT LRR 277.  
 FT LRR 278.  
 FT LRR 279.  
 FT LRR 280.  
 FT LRR 281.  
 FT LRR 282.  
 FT LRR 283.  
 FT LRR 284.  
 FT LRR 285.  
 FT LRR 286.  
 FT LRR 287.  
 FT LRR 288.  
 FT LRR 289.  
 FT LRR 290.  
 FT LRR 291.  
 FT LRR 292.  
 FT LRR 293.  
 FT LRR 294.  
 FT LRR 295.  
 FT LRR 296.  
 FT LRR 297.  
 FT LRR 298.  
 FT LRR 299.  
 FT LRR 300.  
 FT LRR 301.  
 FT LRR 302.  
 FT LRR 303.  
 FT LRR 304.  
 FT LRR 305.  
 FT LRR 306.  
 FT LRR 307.  
 FT LRR 308.  
 FT LRR 309.  
 FT LRR 310.  
 FT LRR 311.  
 FT LRR 312.  
 FT LRR 313.  
 FT LRR 314.  
 FT LRR 315.  
 FT LRR 316.  
 FT LRR 317.  
 FT LRR 318.  
 FT LRR 319.  
 FT LRR 320.  
 FT LRR 321.  
 FT LRR 322.  
 FT LRR 323.  
 FT LRR 324.  
 FT LRR 325.  
 FT LRR 326.  
 FT LRR 327.  
 FT LRR 328.  
 FT LRR 329.  
 FT LRR 330.  
 FT LRR 331.  
 FT LRR 332.  
 FT LRR 333.  
 FT LRR 334.  
 FT LRR 335.  
 FT LRR 336.  
 FT LRR 337.  
 FT LRR 338.  
 FT LRR 339.  
 FT LRR 340.  
 FT LRR 341.  
 FT LRR 342.  
 FT LRR 343.  
 FT LRR 344.  
 FT LRR 345.  
 FT LRR 346.  
 FT LRR 347.  
 FT LRR 348.  
 FT LRR 349.  
 FT LRR 350.  
 FT LRR 351.  
 FT LRR 352.  
 FT LRR 353.  
 FT LRR 354.  
 FT LRR 355.  
 FT LRR 356.  
 FT LRR 357.  
 FT LRR 358.  
 FT LRR 359.  
 FT LRR 360.  
 FT LRR 361.  
 FT LRR 362.  
 FT LRR 363.  
 FT LRR 364.  
 FT LRR 365.  
 FT LRR 366.  
 FT LRR 367.  
 FT LRR 368.  
 FT LRR 369.  
 FT LRR 370.  
 FT LRR 371.  
 FT LRR 372.  
 FT LRR 373.  
 FT LRR 374.  
 FT LRR 375.  
 FT LRR 376.  
 FT LRR 377.  
 FT LRR 378.  
 FT LRR 379.  
 FT LRR 380.  
 FT LRR 381.  
 FT LRR 382.  
 FT LRR 383.  
 FT LRR 384.  
 FT LRR 385.  
 FT LRR 386.  
 FT LRR 387.  
 FT LRR 388.  
 FT LRR 389.  
 FT LRR 390.  
 FT LRR 391.  
 FT LRR 392.  
 FT LRR 393.  
 FT LRR 394.  
 FT LRR 395.  
 FT LRR 396.  
 FT LRR 397.  
 FT LRR 398.  
 FT LRR 399.  
 FT LRR 400.  
 FT LRR 401.  
 FT LRR 402.  
 FT LRR 403.  
 FT LRR 404.  
 FT LRR 405.  
 FT LRR 406.  
 FT LRR 407.  
 FT LRR 408.  
 FT LRR 409.  
 FT LRR 410.  
 FT LRR 411.  
 FT LRR 412.  
 FT LRR 413.  
 FT LRR 414.  
 FT LRR 415.  
 FT LRR 416.  
 FT LRR 417.  
 FT LRR 418.  
 FT LRR 419.  
 FT LRR 420.  
 FT LRR 421.  
 FT LRR 422.  
 FT LRR 423.  
 FT LRR 424.  
 FT LRR 425.  
 FT LRR 426.  
 FT LRR 427.  
 FT LRR 428.  
 FT LRR 429.  
 FT LRR 430.  
 FT LRR 431.  
 FT LRR 432.  
 FT LRR 433.  
 FT LRR 434.  
 FT LRR 435.  
 FT LRR 436.  
 FT LRR 437.  
 FT LRR 438.  
 FT LRR 439.  
 FT LRR 440.  
 FT LRR 441.  
 FT LRR 442.  
 FT LRR 443.  
 FT LRR 444.  
 FT LRR 445.  
 FT LRR 446.  
 FT LRR 447.  
 FT LRR 448.  
 FT LRR 449.  
 FT LRR 450.  
 FT LRR 451.  
 FT LRR 452.  
 FT LRR 453.  
 FT LRR 454.  
 FT LRR 455.  
 FT LRR 456.  
 FT LRR 457.  
 FT LRR 458.  
 FT LRR 459.  
 FT LRR 460.  
 FT LRR 461.  
 FT LRR 462.  
 FT LRR 463.  
 FT LRR 464.  
 FT LRR 465.  
 FT LRR 466.  
 FT LRR 467.  
 FT LRR 468.  
 FT LRR 469.  
 FT LRR 470.  
 FT LRR 471.  
 FT LRR 472.  
 FT LRR 473.  
 FT LRR 474.  
 FT LRR 475.  
 FT LRR 476.  
 FT LRR 477.  
 FT LRR 478.  
 FT LRR 479.  
 FT LRR 480.  
 FT LRR 481.  
 FT LRR 482.  
 FT LRR 483.  
 FT LRR 484.  
 FT LRR 485.  
 FT LRR 486.  
 FT LRR 487.  
 FT LRR 488.  
 FT LRR 489.  
 FT LRR 490.  
 FT LRR 491.  
 FT LRR 492.  
 FT LRR 493.  
 FT LRR 494.  
 FT LRR 495.  
 FT LRR 496.  
 FT LRR 497.  
 FT LRR 498.  
 FT LRR 499.  
 FT LRR 500.  
 FT LRR 501.  
 FT LRR 502.  
 FT LRR 503.  
 FT LRR 504.  
 FT LRR 505.  
 FT LRR 506.  
 FT LRR 507.  
 FT LRR 508.  
 FT LRR 509.  
 FT LRR 510.  
 FT LRR 511.  
 FT LRR 512.  
 FT LRR 513.  
 FT LRR 514.  
 FT LRR 515.  
 FT LRR 516.  
 FT LRR 517.  
 FT LRR 518.  
 FT LRR 519.  
 FT LRR 520.  
 FT LRR 521.  
 FT LRR 522.  
 FT LRR 523.  
 FT LRR 524.  
 FT LRR 525.  
 FT LRR 526.  
 FT LRR 527.  
 FT LRR 528.  
 FT LRR 529.  
 FT LRR 530.  
 FT LRR 531.  
 FT LRR 532.  
 FT LRR 533.  
 FT LRR 534.  
 FT LRR 535.  
 FT LRR 536.  
 FT LRR 537.  
 FT LRR 538.  
 FT LRR 539.  
 FT LRR 540.  
 FT LRR 541.  
 FT LRR 542.  
 FT LRR 543.  
 FT LRR 544.  
 FT LRR 545.  
 FT LRR 546.  
 FT LRR 547.  
 FT LRR 548.  
 FT LRR 549.  
 FT LRR 550.  
 FT LRR 551.  
 FT LRR 552.  
 FT LRR 553.  
 FT LRR 554.  
 FT LRR 555.  
 FT LRR 556.  
 FT LRR 557.  
 FT LRR 558.  
 FT LRR 559.  
 FT LRR 560.  
 FT LRR 561.  
 FT LRR 562.  
 FT LRR 563.  
 FT LRR 564.  
 FT LRR 565.  
 FT LRR 566.  
 FT LRR 567.  
 FT LRR 568.  
 FT LRR 569.  
 FT LRR 570.  
 FT LRR 571.  
 FT LRR 572.  
 FT LRR 573.  
 FT LRR 574.  
 FT LRR 575.  
 FT LRR 576.  
 FT LRR 577.  
 FT LRR 578.  
 FT LRR 579.  
 FT LRR 580.  
 FT LRR 581.  
 FT LRR 582.  
 FT LRR 583.  
 FT LRR 584.  
 FT LRR 585.  
 FT LRR 586.  
 FT LRR 587.  
 FT LRR 588.  
 FT LRR 589.  
 FT LRR 590.  
 FT LRR 591.  
 FT LRR 592.  
 FT LRR 593.  
 FT LRR 594.  
 FT LRR 595.  
 FT LRR 596.  
 FT LRR 597.  
 FT LRR 598.  
 FT LRR 599.  
 FT LRR 600.  
 FT LRR 601.  
 FT LRR 602.  
 FT LRR 603.  
 FT LRR 604.  
 FT LRR 605.  
 FT LRR 606.  
 FT LRR 607.  
 FT LRR 608.  
 FT LRR 609.  
 FT LRR 610.  
 FT LRR 611.  
 FT LRR 612.  
 FT LRR 613.  
 FT LRR 614.  
 FT LRR 615.  
 FT LRR 616.  
 FT LRR 617.  
 FT LRR 618.  
 FT LRR 619.  
 FT LRR 620.  
 FT LRR 621.  
 FT LRR 622.  
 FT LRR 623.  
 FT LRR 624.  
 FT LRR 625.  
 FT LRR 626.  
 FT LRR 627.  
 FT LRR 628.  
 FT LRR 629.  
 FT LRR 630.  
 FT LRR 631.  
 FT LRR 632.  
 FT LRR 633.  
 FT LRR 634.  
 FT LRR 635.  
 FT LRR 636.  
 FT LRR 637.  
 FT LRR 638.  
 FT LRR 639.  
 FT LRR 640.  
 FT LRR 641.  
 FT LRR 642.  
 FT LRR 643.  
 FT LRR 644.  
 FT LRR 645.  
 FT LRR 646.  
 FT LRR 647.  
 FT LRR 648.  
 FT LRR 649.  
 FT LRR 650.  
 FT LRR 651.  
 FT LRR 652.  
 FT LRR 653.  
 FT LRR 654.  
 FT LRR 655.  
 FT LRR 656.  
 FT LRR 657.  
 FT LRR 658.  
 FT LRR 659.  
 FT LRR 660.  
 FT LRR 661.  
 FT LRR 662.  
 FT LRR 663.  
 FT LRR 664.  
 FT LRR 665.  
 FT LRR 666.  
 FT LRR 667.  
 FT LRR 668.  
 FT LRR 669.  
 FT LRR 670.  
 FT LRR 671.  
 FT LRR 672.  
 FT LRR 673.  
 FT LRR 674.  
 FT LRR 675.  
 FT LRR 676.  
 FT LRR 677.  
 FT LRR 678.  
 FT LRR 679.  
 FT LRR 680.  
 FT LRR 681.  
 FT LRR 682.  
 FT LRR 683.  
 FT LRR 684.  
 FT LRR 685.  
 FT LRR 686.  
 FT LRR 687.  
 FT LRR 688.  
 FT LRR 689.  
 FT LRR 690.  
 FT LRR 691.  
 FT LRR 692.  
 FT LRR 693.  
 FT LRR 694.  
 FT LRR 695.  
 FT LRR 696.  
 FT LRR 697.  
 FT LRR 698.  
 FT LRR 699.  
 FT LRR 700.  
 FT LRR 701.  
 FT LRR 702.  
 FT LRR 703.  
 FT LRR 704.  
 FT LRR 705.  
 FT LRR 706.  
 FT LRR 707.  
 FT LRR 708.  
 FT LRR 709.  
 FT LRR 710.  
 FT LRR 711.  
 FT LRR 712.  
 FT LRR 713.  
 FT LRR 714.  
 FT LRR 715.  
 FT LRR 716.  
 FT LRR 717.  
 FT LRR 718.  
 FT LRR 719.  
 FT LRR 720.  
 FT LRR 721.  
 FT LRR 722.  
 FT LRR 723.  
 FT LRR 724.  
 FT LRR 725.  
 FT LRR 726.  
 FT LRR 727.  
 FT LRR 728.  
 FT LRR 729.  
 FT LRR 730.  
 FT LRR 731.  
 FT LRR 732.  
 FT LRR 733.  
 FT LRR 734.  
 FT LRR 735.  
 FT LRR 736.  
 FT LRR 737.  
 FT LRR 738.  
 FT LRR 739.  
 FT LRR 740.  
 FT LRR 741.  
 FT LRR 742.  
 FT LRR 743.  
 FT LRR 744.  
 FT LRR 745.  
 FT LRR 746.  
 FT LRR 747.  
 FT LRR 748.  
 FT LRR 749.  
 FT LRR 750.  
 FT LRR 751.  
 FT LRR 752.  
 FT LRR 753.  
 FT LRR 754.  
 FT LRR 755.  
 FT LRR 756.  
 FT LRR 757.  
 FT LRR 758.  
 FT LRR 759.  
 FT LRR 760.  
 FT LRR 761.  
 FT LRR 762.  
 FT LRR 763.  
 FT LRR 764.  
 FT LRR 765.  
 FT LRR 766.  
 FT LRR 767.  
 FT LRR 768.  
 FT L

Query Match 19.8%; Score 1225.5; DB 1; Length 980;  
 best Local Similarity 29.4%; Pred. No. 2.1e-63;  
 Matches 352; Conservative 189; Mismatches 358; Indels 297; Gaps 35;

FT REPEAT 577 600 LRR 17.  
 FT REPEAT 675 691 LRR 18.  
 FT DOMAIN 692 968 PROTEIN KINASE  
 FT NP BIND 698 706 ATP (BY SIMILARITY)  
 FT BINDING 720 720 ATP (BY SIMILARITY)  
 FT ACT SITE 817 817 BY SIMILARITY  
 FT CARBOHYD 76 76 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 98 98 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 348 348 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 422 422 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 510 510 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 574 574 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 602 602 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CONFLICT 235 235 R -> P (IN REF. 3)  
 FT CONFLICT 703 703 A -> S (IN REF. 2)  
 FT CONFLICT 856 856 G -> D (IN REF. 2)  
 SQ SEQUENCE 980 AA; 107656 MW; B69B2DA7A863295 CRC64;

Query 14 LFFSFFSLFQSPQSGLYREIHQISFQVLT--PDKNLLPDMSSKNF---CTPDGVT 68  
 Db 13 LHTLFFSPCP-----AYTDMEVLLNLTKSSMTGPGHGLHMDIHSSPDHACSFSG 65  
 Qy 69 CRDD-KTSTIDLSKPLNPGSAVSSSLSTGLSESLFLSNHSHGVSQFCKSASLTSL 127  
 Db 66 CDDARVSLNVSFTPL--FGTISPEIGMLTHLVNLTLANNTG-----EL 110  
 Qy 128 DLSRNSLGPVTTLTSLGSCGLKFLNVSNN--TLDPGKVSGLKLNLEVLDSANS 184  
 Db 111 PLEMSL-----TSLKVLNINNNGLTGPGEI----- 139  
 Qy 185 ISGANVGVWVSDGCGELKHLAISNKISGPDVSRCCNLEFLVSSNPFSTGIPFLGDC 244  
 Db 140 -----LKAAMDLEVLDTYNNFN----- 157  
 Qy 245 SALQHLDISGNKLSGDFSRALSTTELKLNINSSNQFVPIPP--LPLKSLQVLSLAENK 302  
 Db 158 -----GKLPEMSELKCLKLKYLSFGGNF 179  
 Qy 303 FTGELIPDLGACDTLTGLDLSGNHFGVAVPPFGSCSLSLALS--SNFSGELPMD-- 359  
 Db 180 FSGELIPESY--GDIQSLLEYGLNAGSGKSPAFSLRLKRLREMYGYNSYTGVPFREG 238  
 Qy 360 TLLKRGKLVLDLSPNFGSELPESLINLSASLLTLDLSSNFGSPILLNCONPKNTLQ 419  
 Db 239 GLTK---LEILIMASCTLTGELIPTLSNN-----KILH 268  
 Qy 420 EYLQNGNGTGKIPPTLSNCSSELSVLSFNYLSGTPSSLSLSKLDKLMMLMGE 479  
 Db 269 TLFILHNNLTGHPPLPSGLVLSKLSLDSINQLTGELIPQSPFLNGLITLINFRNKLVGQ 328  
 Qy 480 IPOELMYVTLETLIDPNDLGEISGLSNCNTNLMWISLNNRLTGLPKMIGLEMLA 539  
 Db 329 IPEALIELPLDLVFEVWENNFTLQLPANLRGNGLIKLVDSDHLLTGLPKOLCGEKLE 388  
 Qy 540 IILKLSNNSPESGNIPELGDORSILIMDLTNLFGNTIPAMFK---QSGKIANFIACK 595  
 Db 389 MLLLSNPFPGPIPEELGKSKSLTKRIYKNLLNGVPLGFLPLVITLILDLNDFSGE 448  
 Qy 596 RYVYIKNDGKKECHGAGNLEFGIARSEQNLSTRNPNCTISRYVGGHTSPPTNNNS 655  
 Db 449 LPTVMSGDVL-----DQVLSNNWFSGEIPPAIGNFBN 481

Qy 656 MMFLDMSYNNLSGYIPKETSMPYLFIINLGHNDISGSIPIDE----- 697  
 Db 482 LQTLFLDRRFRFNIPREIFELHLSRINTSANNITGGIPDSISCRSTLISVDSLRNN 541  
 Qy 698 -----VGLDRGINIDLSNKKDGRIPQMSALTWLREIDLSNNNLSPGIPDMQFET 751  
 Db 542 GEIPKGINNVKNTGLTINISGNQLTGSIPTGIGMNTSLTLDLDFNDLSGRVPLGQFLVF 601  
 Qy 752 PPAKFLNPNGLCGYPLPR---CDPSNADGVANHQRSHGRPPASLAGSVAMGLFSPVCI 808  
 Db 602 NETSFAENTYLC---LPHRVSC--PTRPGQTSDHNTALFSPSRIVTV-----IAALT 650  
 Qy 809 GLTIIVGEMKRRKRKEAELEMYAEGHNSGDRTANNTWKLGVYEAALISINLAPEKEL 868  
 Db 651 GLTILISVALIKOMNKKXOK-----SLAMKLT-----ARQ--- 679  
 Qy 869 RKLTF--ADLLQATNGPHNDLSLGGGFGDVYKALLKQSAVAIKULI--HVSQGGREFM 925  
 Db 680 -KLPFKSEDLVLEC---LKENIIGKGAGIVYRGSMPNNVDVAIKLVGRGTRSDHGF 735  
 Qy 926 AEMETIKIKRINVLPLGYCKYGDDEBLVNEVWKVGSLEEDVQDPKGGVKKLSTRRK 985  
 Db 736 AEIQTGLIRRHRTVRLGVANKDTNLLYETVPSNLSLELHSGSG--HLQWETRRH 793  
 Qy 986 IAIQSARGLAFLHNCSPHIIRDMKSSNYLDDENTLEARYSDFGMARLMSAMDTLSVST 1045  
 Db 794 VAVRAKELCTLHDCSPILIRHRYKSNNTLLDDFPAHVADEFLAKFLVDGAASECMSS 853  
 Qy 1046 LAGTGVVPPPYGOSFRCSYTGVDYSGVTLLELTKRPDSDPDDN--NLVGMQKHA 1104  
 Db 854 IAGSYGYIAPRYAATLAKVDKESDVYSFGVLLTELIAKKVPG--EPEGVDYANVR--NT 910  
 Qy 1105 KLRISDVDPDELMEK--DPALE-----IELQHLKVAAACLDDBARRBPTVQVAM 1154  
 Db 911 EELITQSPDAIIVAVIADPRLTGLPISVHVFRIAMWCVEEAAPRTREYVHM 966

RESULT 3  
 TMKL ARATH STANDARD; FRF; 942 AA.  
 AC P43298;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Putative receptor protein kinase TMKL precursor (EC 2.7.1.1-).  
 GN TMKL OR A11G65150 OR P15E12.4.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC euroside II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_Taxid=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=93076110; PubMed=1332795;  
 RA Chang C., Schaller G.E., Patterson S.E., Kwok S.F.,  
 Meyerowitz E.M., Blecker A.B.;  
 RT "The TMKL gene from Arabidopsis codes for a protein with structural  
 and biochemical characteristics of a receptor protein kinase.";  
 RL Plant Cell 4:1263-1271 (1992).  
 RU [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=21016719; PubMed=11130712;  
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,  
 White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,  
 Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
 Chung M.K., Corn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
 Dunn P., Etgu P., Feildlyum T.V., Feng J.-D., Feng B., Fujii C.Y.,  
 Gail J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
 Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
 Kim C.J., Koo H.L., Kremenskata I., Kurtz D.B., Kwan A., Lam B.,



Db 782 LMELITGRKSLDESGPEESHLVSWFKMYINKKASFKKALDTTID---LDEETLASVH 838  
 Qy 1127 LLOHLKVAVACLDDBAMRRPTM---VQVAMFKEIQAGS-----GID----- 1165  
 Db 839 TVAEI--AGHCARREPQRPDMGHAVNTLSLVELMKPSQNPEDYIGDIDLMSLPQALK 896  
 Qy 1166 -----SQSTIRSIEDGGFSTIEMVMSIKEVEEG 1194  
 Db 897 KMQAVEGRSDLESTSTSLPSLDNTQMSIPTRPYG 931

RESULT 4  
 ID TML1 ARATH STANDARD; PRT; 674 AA.  
 AC P3543;  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 41, Last annotation update)  
 DE Putative kinase-like protein TMKL1 precursor.  
 DE TMKL1 OR AT3G24660 OR MSD24.3.  
 OC Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eustoids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia; TISSUE=Green siliques;  
 RX MEDLINE=9403320; PubMed=8219075;  
 RA Valon C., Smalle J., Goodman H.M., Giraudet J.,  
 RT "Characterization of an Arabidopsis thaliana gene (TMKL1) encoding a  
 RT putative transmembrane protein with an unusual kinase-like domain.",  
 RL Plant Mol. Biol. 23:415-421 (1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=2036309; PubMed=10907853;  
 RA Kaneke T., Katoh T., Sato S., Nakamura A., Asamizu E., Tabata S.,  
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.  
 RT Sequence features of the 4,251,695 bp regions covered by 90 pl. YAC  
 RT and BAC clones".  
 RL DNA Res. 7:217-221 (2000).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: TO PROTEIN KINASES IN THE C-TERMINUS. BUT DOES NOT  
 CC SEEM TO HAVE CONSERVED A KINASE ACTIVITY.  
 CC -1- SIMILARITY: CONTAINS 7 LEUCINE-RICH REPEATS (LRR).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X72863; CAAS1385.1; -  
 DR EMBL: AP000740; BAB01215.1; -  
 DR PIR: S35397; S35397.  
 DR PIR: S39476; S39476.  
 DR InterPro: IPR000719; Euk pkinaase.  
 DR InterPro: IPR001611; LRR.  
 DR InterPro: IPR003592; LRR\_out.  
 DR Pfam: PF00560; pkinaase; 1.  
 DR Pfam: PF00560; LRR; 5.  
 DR ProDom: PD000001; Euk pkinaase; 1.  
 DR SMART: SM00370; LRR; 4.  
 DR PROSITE: PS50011; PROTEIN KINASE DOM; 1.  
 DR Receptor; Transmembrane; Glycoprotein; Signal; Leucine-rich repeat;  
 KW Repeat.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 674 PUTATIVE KINASE-LIKE PROTEIN TMKL1.  
 FT DOMAIN 26 295 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 296 323 POTENTIAL.

FT DOMAIN 324 674 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 373 674 PROTEIN KINASE LIKE.  
 FT REPEAT 98 122 LRR 1.  
 FT REPEAT 123 146 LRR 2.  
 FT REPEAT 147 170 LRR 3.  
 FT REPEAT 172 195 LRR 4.  
 FT REPEAT 198 224 LRR 5.  
 FT REPEAT 226 245 LRR 6.  
 FT REPEAT 246 268 LRR 7.  
 FT CARBOHYD 57 90 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 90 95 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 183 183 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 195 195 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 674 AA; 73353 MW; 5330A8D8767E293 CRC64;

Query Match 10.3%; Score 639; DB 1; Length 674;

Best Local Similarity 27.6%; Pred. No. 1,1e-29; Indels 156; Gaps 25;

Matches 209; Conservative 137; Mismatches 254;

Qy 445 LHLSPNYL-----SGTIPSSIGSLKRDIAKMLNMEGEP--QELMYKTEFTLI-- 494  
 Db 9 LHVIFPVLILHCHCQT-----SLSGSDVKKLLIGKIKSLQGNSSBLSSWNSSVVPV 62  
 Qy 495 -----LDNDLTGE--IPSGLSNCPNINWISLS--NNRLTGEIPKMGRL 535  
 Db 63 COMRGKVFWSNGSPLOQSDSSPQWNTSLFNDSLHLSLQLPANLTSLPREIGEF 122  
 Qy 536 EMUATIKLSNNSPFGNIPDELGDCLSLIWLDTNLTPNGTIPAAFKQSGKIAANFAGK 595  
 Db 123 SMLQSVFLNINLSGSIPLSELGYTSSLDVDSGKALAGVLPpsIWNLCDKL----- 174  
 Qy 596 RYVYIKNDMKKECHGAGNLEFGIRSEQLNRSLSTRPCNITSRYVYGHSPTFDNNGS 655  
 Db 175 -----VSFKIHG----- 181  
 Qy 656 MAFPLMSYMMISGYIPKEIGSWP-----YLFILNCHNDISGIPDEYDGLNLIDLS 710  
 Db 182 -----NNLSGVLEPP--ALPNSTCQNLQVLDGSKNFSEFPEFTIRFGVAGSLDS 231  
 Qy 711 SNKLDRIPOAMSAITMLTLEIDLNNNLISGPIPMGQETPEPPAKFLNPPGLCGPLPRC 770  
 Db 232 SNVFBGLVVEGLGVLE--LSTLNLNHNFGMLPDRGESFGAISEGNSPSCGLPLKPC 290  
 Qy 771 DPSNADGVIAHQSHGRPRASLAGSVAMGLFSPYCIFELLVGRBMRKREAELEM 830  
 Db 291 LGSRR-----LSPGAVAGLV--IGLMSGAVVAVASLLGYLQNKRRKSSISEDDL 338  
 Qy 831 YAGHNSGSGDRNTANNNTNMTLGVKEALSINLAFAFEKPLKTLFPADLLQATNGFHNDLSIG 890  
 Db 339 -----EEGDEE-----IGKEGGBGLVVFQGG--ENLTLDDVLANTG-----QVME 380  
 Qy 891 SGFGDVYKAILKDGSAVAIKKLHVSGQDREPMAMETIGIKIRNVLPLGIC--KV 948  
 Db 381 KTSYGTVYKAKLSDGNNIALRLIREGTCXDRSSCLPIVIRQIGIRHENLVPLRAFOGGR 440  
 Qy 949 GDERLVNEVMYKGSLEEDVYQDPKSGVTLKSTRKIKIGSARGAFPHNCSPIIHR 1008  
 Db 441 G-EKLLIYDYLPNISLHDLHESKPRKPALNWARRKIALGIALKGLAYLHTQGEVYIING 499  
 Qy 1009 DMKSSNVLLDENLEARVSDGMARLM--SAMDTHLSVSTLAGTPGVVPPYQSFRCSTK 1066  
 Db 500 NIKSKNVLVDPFARLTEREGDKIMVQAVAD--EIVSQAKSDGYKAKABELHMKCKNRR 556  
 Qy 1067 GDVYSYGVLLLELTGRRPTDSPDPGDN--NLVGYWKQHA--KLRIDVDPPELMK--EDP 1121  
 Db 557 SDVYAFGLLELTLMGKKPKSGKNGNEFVDLPSTLYKAAVLEETWEVDLEAMKGRSD 616  
 Qy 1122 ALLELLQHLKVAVACLDDBAMRRPTMVQVAMFKE 1157



ALTERED, AND SURFACE FUNCTIONS ARE COMPROMISED, ALLOWING GRAFT-LIKE FUSIONS BETWEEN ORGANS.

- SUBCELLULAR LOCATION: Type 1 membrane protein.

- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

- SIMILARITY: CONTAINS 1 TNFR-CYS REPEAT.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; U67422; AAB09771.1; -

MaizeDB; 128723; -

DR InterPro: IPR000719; Btk\_pkinase.

DR InterPro: IPR004040; STY\_pkinase.

DR InterPro: IPR002290; Ser\_thr\_pkinase.

DR InterPro: IPR001368; TNFR\_c6.

DR Pfam: PF000020; TNFR\_c6; 1.

DR Pfam: PF000069; pkinase; 1.

DR ProDom: PD000001; Btk\_pkinase; 1.

DR SMART: SM00208; TNFR; 1.

DR SMART: SM00107; PROTEIN KINASE\_ATP; 1.

DR PROSITE: PS00108; PROTEIN KINASE\_ST; 1.

DR PROSITE: PS50011; PROTEIN KINASE\_DOM; 1.

DR PROSITE: PS00652; TNFR\_NGFR\_1; FALSE\_NEG.

DR PROSITE: PS50050; TNFR\_NGFR\_2; 1.

DR Receptor, Transferase; Serine/threonine-protein kinase; ATP-binding; Repeat; Transmembrane; Signal.

KW SIGNAL 1 24 POTENTIAL.

FT CHAIN 25 901 PUTATIVE RECEPTOR PROTEIN KINASE

FT TRANS 25 901 CRINKLY4.

FT DOMAIN 25 423 EXTRACELLULAR (POTENTIAL).

FT TRANS 424 444 POTENTIAL.

FT DOMAIN 445 901 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 33 330 7 X 36 AA REPEATS.

FT REPEAT 33 68 1.

FT REPEAT 72 107 2.

FT REPEAT 125 160 3.

FT REPEAT 162 195 4.

FT REPEAT 203 236 5.

FT REPEAT 253 287 6.

FT REPEAT 292 330 7.

FT REPEAT 357 391 TNFR\_CYS.

FT DOMAIN 505 519 PROTEIN KINASE.

FT NP BIND 511 519 ATP (BY SIMILARITY).

FT BINDING 533 533 ATP (BY SIMILARITY).

FT ACT SITE 634 634 BY SIMILARITY.

SEQUENCE 901 AA; 97439 MW; 5258481AC187E061 CRC64;

Query Match 8.5%; Score 528; DB 1; Length 901;

Best Local Similarity 39.8%; Pred. No. 4,3e-23;

Matches 131; Conservative 63; Mismatches 101; Indels 34; Gaps 13;

QY 841 RTANNT-----NWKLTVKKAALSTINLAFAFKLKLFFADLLQATNGFNHDSLGSG 893

DB 462 RLAKSTAVSRKDKMKIQPMEDKLIRRA-----QFSYEIEQKATGFSFSDSOVGKGS 515

QY 894 FGDVYFAILLKDGSAVAIKLIHVS--QGDREPMAMETIGIKIKNRLVPLLGYCKVGD 951

DB 516 FSCVFGKILRDGTVAVAKIKASDVKSKSEPHNLDLSRLNVAHLNLGLGYCEDGSE 575

QY 952 RLIVNEVMKYSLEIVL--ODP--KKGAVLKLISTRKKAISGARGLATLVHNCSPHIIH 1007

DB 576 RLIVYEMARGSLYQHLHGKDPMLKK---RLNWARVTLAVQAKRGIEYHAGYACPPVTH 632

QY 1008 RDKKSSNVLLDEULKARVSPFGMARLMSAMDTLSVTL--AGTPGVVPEYVYQSFCSTK 1066

DB 633 RDIKSSNILLDEHNAKAVDPGLS--ILGPADSGTPISELPAAGLGLYLDPEYRYLHLYLTK 691

QY 1067 GDVYSYGVLLLETLITGRKPRYDSDPGDNNLVGVYKQHKARISDFV---DPELMKEDPAL 1123

DB 692 SDVYSFVLLLETLITSGKALDM-QFEEGNIVEMAV--PLIKAGDIPALDPVL---SPSS 745

QY 1124 EIELLOHL-KVAVACLDLDRMRPTWQV 1151

DB 746 DLEALKKIASVACVCKVEMRGKDRPSMDKV 774

RESULT 7

APKA ARATH STANDARD; PRT; 410 AA.

ID APKA ARATH

AC Q06548;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Protein Kinase APKA (EC 2.7.1.1-).

GN APKA.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; Eusteroideae; Brassicales; Brassicaceae; Arabidopsis.

OC NCBI\_Taxid=3702;

OK NCBI\_Taxid=3702;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=93081726; PubMed=1450380;

RA Hirayama T., Oka A.;

RT "Novel protein kinase of Arabidopsis thaliana (APKA) that phosphorylates tyrosine, serine and threonine."

RL Plant Mol. Biol. 20:653-662(1992).

RT phosphotyrosine kinase.

CC - FUNCTION: POSSIBLE BI-FUNCTIONAL KINASE. IN VITRO, IT EXHIBITS SERINE/THREONINE ACTIVITY. IN VIVO, CAN PHOSPHORYLATE TYROSINE RESIDUES OF LIMITED SUBSTRATES.

CC - SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; D12522; BAA02092.1; -

DR InterPro: IPR000719; Btk\_pkinase.

DR InterPro: IPR004040; STY\_pkinase.

DR InterPro: IPR002290; Ser\_thr\_pkinase.

DR Pfam: PF000069; pkinase; 1.

DR ProDom: PD000001; Btk\_pkinase; 1.

DR SMART: SM00221; STYKc; 1.

DR PROSITE: PS00108; PROTEIN KINASE\_ATP; 1.

DR PROSITE: PS00107; PROTEIN KINASE\_ST; 1.

DR PROSITE: PS50011; PROTEIN KINASE\_DOM; 1.

DR Transferase; Serine/threonine-protein kinase; Tyrosine-protein kinase; ATP-binding; Multigene family; Myristate.

KW ATP-binding; Multigene family; Myristate.

FT LIPID 2 MYRISTATE (BY SIMILARITY).

FT DOMAIN 68 352 PROTEIN KINASE.

FT NP BIND 74 82 ATP (BY SIMILARITY).

FT BINDING 106 106 ATP (BY SIMILARITY).

FT ACT SITE 203 203 BY SIMILARITY.

SEQUENCE 410 AA; 45519 MW; 5BAB28D9E0065082 CRC64;

Query Match 8.2%; Score 508; DB 1; Length 410;

Best Local Similarity 36.5%; Pred. No. 2.1e-22;

Matches 129; Conservative 61; Mismatches 131; Indels 32; Gaps 9;

QY 868 LRKLPADLLQATNGFNHDSLGSGGDVYKALILD-----GSAVAIKLIHVS 917

DB 53 LKSFSEALKSATNFPDSDVLSGEGFCYFKGMIDKSLTASRPGGLVAVKXLNQD 112

QY 918 GQGDREPMAMETIGIKIKNRLVPLLGYCYGVGERLLVNEVMKYSLEIVLQDPKKGAVK 977

Db 113 MGHQEWLAENVYLGQFSHRLVLYCLDEHRLLYVEFMPSGLIENHLLFRGGLYFOP 172  
 Qy 978 LKSTRKRLAIGSARGALFLHNSCPHIIHRDMKSNVLLDENLEAVSPDGMARLMSAM 1037  
 Db 173 LSWRLKRLVLAAGAKGALFL-HSETRVLYRDPFKTSNILLDSYNAKLSDFGLAKGPGIG 231  
 Qy 1038 D-THLSVSTLAGEGYVPEYVYOSFRSTGVDVYVYVVLLELTGRPTD-SPDFGDN 1095  
 Db 232 DKSHVSTRVM-GTIGYVAPFYLATGHLTTSDVYSFGVVLLELTSGRRVADKRPFGERN 290  
 Qy 1096 LVGVKQHA-KRLSDVFPPELMKEDPALEIELQHLKVAVACLDRAWRPPTVQVNA 1153  
 Db 291 LVEMAKFYLVRKRIFFRVIDNRL--QDQYMEBACVATLISRLCTTEIKAPNMSSEVS 348  
 Qy 1154 MFKEIOA-----GSGID-----SQTIRSIEDGGFSTTEMVMSIKEVP 1192  
 Db 349 HLEHTQSLNAIGGNMDKTRDRMRSDSVSKVNAAGFARQAVGSTVAVP 401

REMARK 8  
 ARATH  
 IAPKB ARATH STANDARD; PRT; 412 AA.  
 AC P46573; 09SLHS;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Protein kinase APK1B (EC 2.7.1.-).  
 GN APK1B OR AT2G28930 OR T914.1.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Rensing C.M., Koo H.L.,  
 RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,  
 RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,  
 RA Goodman H.M., Somerville C.R., Copenhagen G.P., Preuss D.,  
 RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Frazer C.M.,  
 RA Venter J.C.;  
 RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 thaliana.";  
 RT Nature 402:761-768(1999).  
 RL [12]  
 RC SEQUENCE OF 143-346 FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA MEDLINE=93081726; PubMed=1450380;  
 RA Hirayama T., Oka A.;  
 RA "Novel protein kinase of Arabidopsis thaliana (APK1) that  
 RT phosphorylates tyrosine, serine and threonine.";  
 RL Plant Mol. Biol. 20:653-662(1992).  
 CC -1- FUNCTION: POSSIBLE BI-FUNCTIONAL KINASE. IN VITRO, IT EXHIBITS  
 CC SERINE/THREONINE ACTIVITY. IN VIVO, CAN PHOSPHORYLATE TYROSINE  
 CC RESIDUES OF LIMITED SUBSTRATES (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AC005315; AAC33221.1; -  
 DR EMBL; D10152; BAA20968.1; -  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR004040; STY\_pkinase.

DR InterPro; IPR002290; Ser\_thr\_kinase.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00669; pkinase; 1.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00221; STYK; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
 KW Transferrase, Serine/threonine-protein kinase; Tyrosine-protein kinase;  
 KW ATP-binding; Multigene family; Myristate.  
 FT IPIID 2 2 MYRISTATE (BY SIMILARITY).  
 FT DOMAIN 69 356 PROTEIN KINASE.  
 FT NP BIND 75 83 ATP (BY SIMILARITY).  
 FT BINDING 107 107 ATP (BY SIMILARITY).  
 FT ACT SITE 204 204 BY SIMILARITY.  
 SQ SEQUENCE 412 AA; 45746 MW; EBICAOB1A626A5DA CRC64;

Query Match 8.2%; Score 507.5; DB 1; Length 412;  
 Best Local Similarity 36.2%; Pred. No. 2.2e-22;  
 Matches 138; Conservative 68; Mismatches 126; Indels 49; Gaps 15;

Qy 830 MTAEGHNSGDRTPAN--NTMVKLTGVKEAL-SINLAFFKPLRKTFFADLQATNGFEHN 885  
 Db 21 MSESANDSLGSKSSVSIRTPRTGG--EILQSPNLSF-----TFALKATNFRP 71  
 Qy 886 DSLGSGGFGVDYKALIND-----GSAVAIKLIHVSOGGDEPFMAENETGKIK 935  
 Db 72 DSVLGGGGGVSFKKMDIEQTLTKSPGTGVIAVKTLMDQMGQHEWLAENVYLGQFS 131  
 Qy 936 HNNVPLLYGCVKGERLLVNVKYGSLSDVLDOPKKGAV--ELKSTRKRLAIGSARG 993  
 Db 132 HNNVPLLYGCVKGERLLVNVKYGSLSDVLDOPKKGAV--ELKSTRKRLAIGSARG 189  
 Qy 994 IAFIHHNSCPHIIHRDMKSNVLLDENLEAVSPDGMARLMSAM--THLSVSTLAGEPY 1052  
 Db 190 IAFI-HNSETSVLYRDPFKTSNILLDSYNAKLSDFGLAKGPGDKSHVS-TIMGTYYG 247  
 Qy 1053 VPEYVYOSFRSTGVDVYVYVVLLELTGRPTD--PPFGDNVLGVKQ--HAKLRI 1108  
 Db 248 AAFETLALGHLTTSDVYSFGVVLLELTSGRRVADKRP--GQKLVEMARPLANKRL 306  
 Qy 1109 SDVPPELMKEDPALEIELQHLKVAVACLDRAWRPPTVQVNAFMFKEIOAGSGIDSQ 1168  
 Db 307 FRVIDNRL--QDQYMEBACVATLISRLCTFEIKAPNMSSEVSHLEHQ----- 355  
 Qy 1169 TIRSIEDGGFSTTEMVMSIKEVP 1192  
 Db 356 ---TINAGGRNIDVQRRWR 373

RESULT 9  
 XPRO MAIZE STANDARD; PRT; 817 AA.  
 ID XPRO MAIZE  
 AC P17801;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Putative receptor protein kinase ZMPK1 precursor (EC 2.7.1.37).  
 GN PK1.  
 OS Zea mays (Maize).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
 OC Panicoideae; Andropogoneae; Zea.  
 OX NCBI\_TaxID=4577;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. B73; TISSUE=Root;  
 RC MEDLINE=90294911; PubMed=2163028;  
 RA Walker J.C., Zhang R.;  
 RA "Relationship of a putative receptor protein kinase from maize to the  
 RT S-locus glycoproteins of Brassica.";  
 RL Nature 345:743-746(1990).



[2] SEQUENCE FROM N.A.  
 RC STRAIN=cv. B73;  
 RA Zhang R., Walker J.C.,  
 RL Submitted (Oct-1992) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: PROBABLE RECEPTOR. INTERACTION WITH A LIGAND IN THE  
 CC EXTRACELLULAR DOMAIN TRIGGERS THE PROTEIN KINASE ACTIVITY OF THE  
 CC CYTOPLASMIC DOMAIN.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. THE SHOOTS AND  
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE SHOOTS AND  
 CC ROOTS OF YOUNG MAIZE SEEDLINGS, AND TO A LESSER EXTENT IN THE  
 CC SILKS.  
 CC -1- SIMILARITY: THE EXTRACELLULAR DOMAIN IS SIMILAR TO S-LOCUS  
 CC GLYCOPROTEINS OF BRASSICA, WHILE THE INTRACELLULAR DOMAIN IS  
 CC A SER/THR-PROTEIN KINASE CLOSELY RELATED TO RAF KINASES.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X52384; CA36611.1; -  
 DR EMBL: X67733; CA47962.1; -  
 DR MaizeDB: 65910;  
 DR InterPro: IPR001480; B\_lectin.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR003014; PAN.  
 DR InterPro: IPR003609; Pan\_app.  
 DR InterPro: IPR002290; Ser\_thr\_pkinase.  
 DR InterPro: IPR000858; Slocus\_glycop.  
 DR Pfam: PF00024; PAN; 1.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Pfam: PF00954; S\_locus\_glycop; 1.  
 DR Pfam: PF01453; Agglutinin; 1.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00108; B\_lectin; 1.  
 DR SMART: SM00181; EGF; 1.  
 DR SMART: SM00473; PAN\_Ap; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE: PS00111; PROTEIN\_KINASE\_DOM; 1.  
 DR Transferrase; Serine/threonine-protein kinase; ATP-binding;  
 DR Transmembrane; Receptor; Glycoprotein; Signal.  
 DR SIGNAL 1 28  
 DR CHAIN 1 28  
 DR DOMAIN 29 472 PUTATIVE RECEPTOR PROTEIN KINASE ZMPK1.  
 DR TRANSMM 473 498 EXTRACELLULAR (POTENTIAL).  
 DR DOMAIN 499 817 CYTOPLASMIC (POTENTIAL).  
 DR DOMAIN 534 817 PROTEIN KINASE.  
 DR NP\_BIND 540 548 ATP (BY SIMILARITY).  
 DR BINDING 562 562 ATP (BY SIMILARITY).  
 DR ACT\_SITE 658 658 BY SIMILARITY.  
 DR CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 DR CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 DR CARBOHYD 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 DR CARBOHYD 279 279 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 DR CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 DR CARBOHYD 339 339 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 DR CARBOHYD 452 452 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 DR SEQUENCE 817 AA; 91120 MW; F164B4471922E67 CRC64;

Query Match 8.2%; Score 507; DB 1; Length 817;  
 Best Local Similarity 26.7%; Pred. No. 6,2e-22;  
 Matches 187; Conservative 113; Mismatches 236; Indels 164; Gaps 30;

QY 531 WIGRLNTALTLKLSNFSGNIDELDCRSLIWLDTNLTFLN-----GITPAMAFKQS 584  
 DB 188 YIRFSPDLSTLSIY-----HVP-QVSD---IYWPDPDQTLYODGRNQYNSTRGLMTDS 238

QY 585 GKIA-ANFLAGKRYIYIK-NDGMKKE---CHGAGNLLFEQGIKRE---OLNRLSTENPCNI 637  
 DB 239 GVLASDSDFDGQALVADVGQVYKRRLLTDPDGNLRILYSANDSDSMGVSVMAMPQNCI 298  
 QY 638 TSHRYVG-----GHTSPFEDNNGSMWFLDMSYNNLSGYIPREISMPYFLPLINGHDI 690  
 DB 299 ----HGLCGPNGICHSPF-----PFCSCPQYATRNPGNW-----TEGCMAI 337  
 QY 691 SGSIPEVEGDLRGNTLIDLSNKLDRIRPQMSALTMFL--RILDSNNLSGPPIEMGQF 748  
 DB 338 VNTTCRY-DKRSRRFVRLPNTDFWGSDDQHLSTLRTCRDICTSDCKCFQYQEGG 396  
 QY 749 ETTPPAFLANPGLCGP-----LPRCD-----PSNADGYAHQ 782  
 DB 397 SCYPAKALFSGR---TYFSDVPTIYKLPLTQSVSNALIPRSDVDSPRRLD--CDRM 451  
 QY 783 RSHGRAP-----ASLAGSVAMGLFSPVCI FGLIIVG-----REMRRRRKKA 826  
 DB 452 NKSIRFPPVHKVKGSGESKWFYGFIAFPVVEVSFISFAWFVYKELRPS----- 505  
 QY 827 ELEMYAGHGNSGDRITANNNTWKLTGVKEALSINLAFEKPLKLTFFADLQATNGFHD 886  
 DB 506 --ELMASEKG-----YKMTSN-----FRYGRRLVYKTRPKVE 539  
 QY 887 SLISGGFGDYKAILDKGSAVAIKKLIHVSQGGREFEAMETTGKIKRNLVPLGYC 946  
 DB 540 --LGRGSGTYTKVGLDDRDHVAVKLENYR--QKVEFQALSVIGGINMNLVRIWFGC 596  
 QY 947 KYGDERLLVNEWYKGLSDYLDQPKKGQYKLTSTRKIAISARGALFLHNCSPHI 1006  
 DB 597 SEGSHRLIVSEYVENSGSLANTLIFS--EGGNILLDWEGFNLALGVAKALVYHHECLEWI 655  
 QY 1007 HRDKSSNVLLDENLEAVSDFGMARLMSAMDTHLSYSLAGTPGYVPEYQSFRCSTK 1066  
 DB 656 HDVAPENILLDQAFEPKITPFGVLKLNREGSQTQNSHVGTLGYLAPEWSSLPITAK 715  
 QY 1067 GDVSYGVLLLELTGKRPDTPSPFGDNNLVGWYQ--HAKLR-----I 1108  
 DB 716 VVYSYGVLLLELTGTRVSE-----LVGQDEHVSMLRKLVRLMSAKLBSBBSWI 767  
 QY 1109 SVVPEPMLKEDPALEIFELLOKLVAVACLDRAWRARPDM 1148  
 DB 768 DQYLSKILNR--PVNYQARTLILKLAVSCLEBDSKRTM 805

RESULT 10  
 SRK6\_BRAOL STANDARD; PRT; 849 AA.  
 AC 009092;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Putative serine/threonine kinase receptor precursor (EC 2.7.1.37)  
 DE (S-receptor kinase) (SRK).  
 GN SRK6.  
 OS Brassica oleracea (Cauliflower).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Brassica.  
 OX NCBI\_TaxID=3712;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. 56S6; TISSUE=Stigma;  
 RX MEDLINE=92020942; PubMed=1681543;  
 RA Stein J.C., Howlett B., Boyes D.C., Nasrallah M.E.;  
 RT "Molecular cloning of a putative receptor protein kinase gene encoded  
 RT at the self-incompatibility locus of Brassica oleracea.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:8816-8820(1991).  
 CC -1- FUNCTION: INVOLVED IN SPOROPHYTIC SELF-INCOMPATIBILITY SYSTEM  
 CC (THE INABILITY OF FLOWERING PLANTS TO COMBINATION WITH S-LOCUS-  
 CC FERTILIZATION), PROBABLY ACTING IN COMBINATION WITH A LIGAND IN THE  
 CC SPECIFIC GLYCOPROTEINS. INTERACTION WITH A LIGAND IN THE



CC	EMBL; X66482; CAA47109.2; ALT_PFRAME.
DR	APO02044; BAB02252.1; -
DR	EMBL; AC069474; AAG51016.1; -
DR	EMBL; AY057652; AAL15283.1; -
DR	PIR; A46260; A46260.
DR	PIR; S22863; S22863.
DR	InterPro; IPR001611; LRR_out.
DR	InterPro; IPR003592; LRR_out.
DR	Pfam; PF00560; LRR; 3.
DR	SMART; SM00370; LRR; 8.
KW	Leucine-rich repeat; Repeat; DNA repair; DNA damage; Signal.
FT	SIGNAL 1 26 POTENTIAL.
FT	CHAIN 27 372 DNA-DAMAGE-REPAIR/TOLERATION PROTEIN
FT	FT REPEAT 134 158 LRR 1.
FT	FT REPEAT 160 182 LRR 2.
FT	FT REPEAT 183 205 LRR 3.
FT	FT REPEAT 207 230 LRR 4.
FT	FT REPEAT 231 256 LRR 5.
FT	FT REPEAT 258 278 LRR 6.
FT	FT REPEAT 279 301 LRR 7.
FT	FT REPEAT 302 326 LRR 8.
FT	FT REPEAT 328 350 LRR 9.
SO	SEQUENCE 372 AA; 39556 MW; BA4361101AC45659 CRCC4;
Query Match	6.5%; Score 401.5; DB 1; Length 372;
Best Local Similarity	28.9%; Pred.No. 2.6e-16;
Matches 116; Conservative 45; Mismatches 133; Indels 107; Gaps	
Oy	368 KVLDSFNEFSGELPESLNLASLLTLDLSNNFSGPLNLCQNPNTQLQELYLN-N 426
Dd	74 RVTDISLR---GESEDDAIPKAG-----RSGYMSGSDPAVCD--LTALTSLVLADWK 121
Oy	427 GTTGKIPPTLSNGEELSYLSHFNYSSTGPSGLSKRDPLKLMLMGEIPELMY 486
Dd	122 GITGGIPCTISLASLIIDLAKNKITGEIPAIGKLSCLAVLAENQMGEIPASITS 181
Oy	487 VTTLETLILDNDITGEIPSGNSCTNLNMVISNNRLTGEIPKIGGLEMLAILIKSNN 546
Dd	182 ILEIMHLETENGITGVIPADFGSLKMLSHVLLGRNELGSIPESISGMERLADDLSKN 241
Oy	547 SSSGNIPDELDCSLIMDLINTLFNGTLPAMFFQSGLIANFIAGRVYVIKNDGW 606
Dd	242 HIEGPLPEWMGMYKVLSDLDCNSLTGPSPGISLNSGDVAAN----- 285
Oy	607 KECHGAGNILEFOGIRPOLNRILSTRNPCNITSRYVGHTSPFPDNGSAMFLDMSYML 666
Dd	286 -----LSRNAL 291
Oy	667 SGYIPEKETSMPYLFIINLGHNDISIPPVGDPLRGNIILDLSSNKLDGRIPQMSALT 726
Dd	292 EGTIPVPFVSFKTYLVSLDSLHSNLSGRIPDLSAKRVGHLDISHKKLGRIPTGF---- 347
Oy	727 MLTEIDLSNNLSGPIPEWGQFETFPAPKFIANNPGICGYPL 767
Dd	348 -----PFDHEATSFEDNQCLCGGPL 368
RESULT 12	
SHO2_HUMAN	STANDARD; PRT; 582 AA.
ID	_SHO2_HUMAN
AC	09U013; O76063;
DT	16-OCT-2001 (Rel. 40; Created)
DT	16-OCT-2001 (Rel. 40; Last sequence update)
DT	15-JUN-2002 (Rel. 41; Last annotation update)
DE	Leucine-rich repeat protein SHO2-2 (Ras-binding protein Sur-8).
GN	SHO2 OR KIA0862.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Methylia; Primates; Catarrhini; Hominiidae; Homo.
OX	NCBI_TaxId=9606;
RN	[1]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=98284030; PubMed=9618511;  
 RA Sellors L.M., Schutzman J.L., Borland C.Z., Stern M.J.;  
 RT "Soc-2 encodes a leucine-rich repeat protein implicated in fibroblast  
 growth factor receptor signaling";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:6903-6908(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98337190; PubMed=9674433;  
 RA Steurich D.S., Sun Q., Han M.;  
 RT "Soc-8, a conserved Ras-binding protein with leucine-rich repeats,  
 positively regulates Ras-mediated signaling in C. elegans";  
 RL Cell 94:119-130(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=99156230; PubMed=10048485;  
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,  
 Miyajima N., Tanaka A., Kotani H., Nomura N., Ohta O.;  
 RT Prediction of the coding sequences of unidentified human genes.  
 R1. The complete sequences of 100 new cDNA clones from brain which  
 code for large proteins in vitro.;  
 RL DNA Res. 5:355-364(1998).  
 CC -1 SUBUNIT: SPECIFICALLY BINDS K-RAS AND N-RAS BUT NOT H-RAS IN  
 CC VITRO.  
 CC -1 SIMILARITY: CONTAINS 20 LEUCINE-RICH REPEATS (LRR).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed, usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@sib-sib.ch).  
 CC -----  
 DR EMBL: AF054828; AAC25698.1; -;  
 DR EMBL: AF068820; AAC39856.1; -;  
 DR EMBL: AB020669; BAA74885.1; -;  
 DR Genew; HGNC:15454; SHOC2.  
 DR MIM; 602775; -;  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR003591; LRR\_typ.  
 DR Pfam; PF00560; LRR\_17.  
 DR PRINTS; PRO00019; LEURICHRPT.  
 DR SMART; SM00369; LRR\_TYP; 12.  
 KW Repeat; Leucine-rich repeat.  
 KW Repeat; Leucine-rich repeat.  
 FT REPEAT 99 122 LRR 1.  
 FT REPEAT 123 145 LRR 2.  
 FT REPEAT 146 168 LRR 3.  
 FT REPEAT 169 191 LRR 4.  
 FT REPEAT 193 214 LRR 5.  
 FT REPEAT 215 237 LRR 6.  
 FT REPEAT 239 260 LRR 7.  
 FT REPEAT 262 283 LRR 8.  
 FT REPEAT 284 306 LRR 9.  
 FT REPEAT 307 329 LRR 10.  
 FT REPEAT 331 353 LRR 11.  
 FT REPEAT 354 377 LRR 12.  
 FT REPEAT 379 400 LRR 13.  
 FT REPEAT 401 424 LRR 14.  
 FT REPEAT 425 447 LRR 15.  
 FT REPEAT 448 471 LRR 16.  
 FT REPEAT 473 493 LRR 17.  
 FT REPEAT 494 516 LRR 18.  
 FT REPEAT 518 540 LRR 19.  
 FT REPEAT 541 563 LRR 20.  
 FT REPEAT 563 582 LRR 21.  
 FT REPEAT 582 604 LRR 22.  
 FT REPEAT 604 626 LRR 23.  
 FT REPEAT 626 648 LRR 24.  
 FT REPEAT 648 670 LRR 25.  
 FT REPEAT 670 692 LRR 26.  
 FT REPEAT 692 714 LRR 27.  
 FT REPEAT 714 736 LRR 28.  
 FT REPEAT 736 758 LRR 29.  
 FT REPEAT 758 780 LRR 30.  
 FT REPEAT 780 802 LRR 31.  
 FT REPEAT 802 824 LRR 32.  
 FT REPEAT 824 846 LRR 33.  
 FT REPEAT 846 868 LRR 34.  
 FT REPEAT 868 890 LRR 35.  
 FT REPEAT 890 912 LRR 36.  
 FT REPEAT 912 934 LRR 37.  
 FT REPEAT 934 956 LRR 38.  
 FT REPEAT 956 978 LRR 39.  
 FT REPEAT 978 1000 LRR 40.  
 FT REPEAT 1000 1022 LRR 41.  
 FT REPEAT 1022 1044 LRR 42.  
 FT REPEAT 1044 1066 LRR 43.  
 FT REPEAT 1066 1088 LRR 44.  
 FT REPEAT 1088 1110 LRR 45.  
 FT REPEAT 1110 1132 LRR 46.  
 FT REPEAT 1132 1154 LRR 47.  
 FT REPEAT 1154 1176 LRR 48.  
 FT REPEAT 1176 1198 LRR 49.  
 FT REPEAT 1198 1220 LRR 50.  
 FT REPEAT 1220 1242 LRR 51.  
 FT REPEAT 1242 1264 LRR 52.  
 FT REPEAT 1264 1286 LRR 53.  
 FT REPEAT 1286 1308 LRR 54.  
 FT REPEAT 1308 1330 LRR 55.  
 FT REPEAT 1330 1352 LRR 56.  
 FT REPEAT 1352 1374 LRR 57.  
 FT REPEAT 1374 1396 LRR 58.  
 FT REPEAT 1396 1418 LRR 59.  
 FT REPEAT 1418 1440 LRR 60.  
 FT REPEAT 1440 1462 LRR 61.  
 FT REPEAT 1462 1484 LRR 62.  
 FT REPEAT 1484 1506 LRR 63.  
 FT REPEAT 1506 1528 LRR 64.  
 FT REPEAT 1528 1550 LRR 65.  
 FT REPEAT 1550 1572 LRR 66.  
 FT REPEAT 1572 1594 LRR 67.  
 FT REPEAT 1594 1616 LRR 68.  
 FT REPEAT 1616 1638 LRR 69.  
 FT REPEAT 1638 1660 LRR 70.  
 FT REPEAT 1660 1682 LRR 71.  
 FT REPEAT 1682 1704 LRR 72.  
 FT REPEAT 1704 1726 LRR 73.  
 FT REPEAT 1726 1748 LRR 74.  
 FT REPEAT 1748 1770 LRR 75.  
 FT REPEAT 1770 1792 LRR 76.  
 FT REPEAT 1792 1814 LRR 77.  
 FT REPEAT 1814 1836 LRR 78.  
 FT REPEAT 1836 1858 LRR 79.  
 FT REPEAT 1858 1880 LRR 80.  
 FT REPEAT 1880 1902 LRR 81.  
 FT REPEAT 1902 1924 LRR 82.  
 FT REPEAT 1924 1946 LRR 83.  
 FT REPEAT 1946 1968 LRR 84.  
 FT REPEAT 1968 1990 LRR 85.  
 FT REPEAT 1990 2012 LRR 86.  
 FT REPEAT 2012 2034 LRR 87.  
 FT REPEAT 2034 2056 LRR 88.  
 FT REPEAT 2056 2078 LRR 89.  
 FT REPEAT 2078 2100 LRR 90.  
 FT REPEAT 2100 2122 LRR 91.  
 FT REPEAT 2122 2144 LRR 92.  
 FT REPEAT 2144 2166 LRR 93.  
 FT REPEAT 2166 2188 LRR 94.  
 FT REPEAT 2188 2210 LRR 95.  
 FT REPEAT 2210 2232 LRR 96.  
 FT REPEAT 2232 2254 LRR 97.  
 FT REPEAT 2254 2276 LRR 98.  
 FT REPEAT 2276 2298 LRR 99.  
 FT REPEAT 2298 2320 LRR 100.  
 FT REPEAT 2320 2342 LRR 101.  
 FT REPEAT 2342 2364 LRR 102.  
 FT REPEAT 2364 2386 LRR 103.  
 FT REPEAT 2386 2408 LRR 104.  
 FT REPEAT 2408 2430 LRR 105.  
 FT REPEAT 2430 2452 LRR 106.  
 FT REPEAT 2452 2474 LRR 107.  
 FT REPEAT 2474 2496 LRR 108.  
 FT REPEAT 2496 2518 LRR 109.  
 FT REPEAT 2518 2540 LRR 110.  
 FT REPEAT 2540 2562 LRR 111.  
 FT REPEAT 2562 2584 LRR 112.  
 FT REPEAT 2584 2606 LRR 113.  
 FT REPEAT 2606 2628 LRR 114.  
 FT REPEAT 2628 2650 LRR 115.  
 FT REPEAT 2650 2672 LRR 116.  
 FT REPEAT 2672 2694 LRR 117.  
 FT REPEAT 2694 2716 LRR 118.  
 FT REPEAT 2716 2738 LRR 119.  
 FT REPEAT 2738 2760 LRR 120.  
 FT REPEAT 2760 2782 LRR 121.  
 FT REPEAT 2782 2804 LRR 122.  
 FT REPEAT 2804 2826 LRR 123.  
 FT REPEAT 2826 2848 LRR 124.  
 FT REPEAT 2848 2870 LRR 125.  
 FT REPEAT 2870 2892 LRR 126.  
 FT REPEAT 2892 2914 LRR 127.  
 FT REPEAT 2914 2936 LRR 128.  
 FT REPEAT 2936 2958 LRR 129.  
 FT REPEAT 2958 2980 LRR 130.  
 FT REPEAT 2980 3002 LRR 131.  
 FT REPEAT 3002 3024 LRR 132.  
 FT REPEAT 3024 3046 LRR 133.  
 FT REPEAT 3046 3068 LRR 134.  
 FT REPEAT 3068 3090 LRR 135.  
 FT REPEAT 3090 3112 LRR 136.  
 FT REPEAT 3112 3134 LRR 137.  
 FT REPEAT 3134 3156 LRR 138.  
 FT REPEAT 3156 3178 LRR 139.  
 FT REPEAT 3178 3200 LRR 140.  
 FT REPEAT 3200 3222 LRR 141.  
 FT REPEAT 3222 3244 LRR 142.  
 FT REPEAT 3244 3266 LRR 143.  
 FT REPEAT 3266 3288 LRR 144.  
 FT REPEAT 3288 3310 LRR 145.  
 FT REPEAT 3310 3332 LRR 146.  
 FT REPEAT 3332 3354 LRR 147.  
 FT REPEAT 3354 3376 LRR 148.  
 FT REPEAT 3376 3398 LRR 149.  
 FT REPEAT 3398 3420 LRR 150.  
 FT REPEAT 3420 3442 LRR 151.  
 FT REPEAT 3442 3464 LRR 152.  
 FT REPEAT 3464 3486 LRR 153.  
 FT REPEAT 3486 3508 LRR 154.  
 FT REPEAT 3508 3530 LRR 155.  
 FT REPEAT 3530 3552 LRR 156.  
 FT REPEAT 3552 3574 LRR 157.  
 FT REPEAT 3574 3596 LRR 158.  
 FT REPEAT 3596 3618 LRR 159.  
 FT REPEAT 3618 3640 LRR 160.  
 FT REPEAT 3640 3662 LRR 161.  
 FT REPEAT 3662 3684 LRR 162.  
 FT REPEAT 3684 3706 LRR 163.  
 FT REPEAT 3706 3728 LRR 164.  
 FT REPEAT 3728 3750 LRR 165.  
 FT REPEAT 3750 3772 LRR 166.  
 FT REPEAT 3772 3794 LRR 167.  
 FT REPEAT 3794 3816 LRR 168.  
 FT REPEAT 3816 3838 LRR 169.  
 FT REPEAT 3838 3860 LRR 170.  
 FT REPEAT 3860 3882 LRR 171.  
 FT REPEAT 3882 3904 LRR 172.  
 FT REPEAT 3904 3926 LRR 173.  
 FT REPEAT 3926 3948 LRR 174.  
 FT REPEAT 3948 3970 LRR 175.  
 FT REPEAT 3970 3992 LRR 176.  
 FT REPEAT 3992 4014 LRR 177.  
 FT REPEAT 4014 4036 LRR 178.  
 FT REPEAT 4036 4058 LRR 179.  
 FT REPEAT 4058 4080 LRR 180.  
 FT REPEAT 4080 4102 LRR 181.  
 FT REPEAT 4102 4124 LRR 182.  
 FT REPEAT 4124 4146 LRR 183.  
 FT REPEAT 4146 4168 LRR 184.  
 FT REPEAT 4168 4190 LRR 185.  
 FT REPEAT 4190 4212 LRR 186.  
 FT REPEAT 4212 4234 LRR 187.  
 FT REPEAT 4234 4256 LRR 188.  
 FT REPEAT 4256 4278 LRR 189.  
 FT REPEAT 4278 4300 LRR 190.  
 FT REPEAT 4300 4322 LRR 191.  
 FT REPEAT 4322 4344 LRR 192.  
 FT REPEAT 4344 4366 LRR 193.  
 FT REPEAT 4366 4388 LRR 194.  
 FT REPEAT 4388 4410 LRR 195.  
 FT REPEAT 4410 4432 LRR 196.  
 FT REPEAT 4432 4454 LRR 197.  
 FT REPEAT 4454 4476 LRR 198.  
 FT REPEAT 4476 4498 LRR 199.  
 FT REPEAT 4498 4520 LRR 200.  
 FT REPEAT 4520 4542 LRR 201.  
 FT REPEAT 4542 4564 LRR 202.  
 FT REPEAT 4564 4586 LRR 203.  
 FT REPEAT 4586 4608 LRR 204.  
 FT REPEAT 4608 4630 LRR 205.  
 FT REPEAT 4630 4652 LRR 206.  
 FT REPEAT 4652 4674 LRR 207.  
 FT REPEAT 4674 4696 LRR 208.  
 FT REPEAT 4696 4718 LRR 209.  
 FT REPEAT 4718 4740 LRR 210.  
 FT REPEAT 4740 4762 LRR 211.  
 FT REPEAT 4762 4784 LRR 212.  
 FT REPEAT 4784 4806 LRR 213.  
 FT REPEAT 4806 4828 LRR 214.  
 FT REPEAT 4828 4850 LRR 215.  
 FT REPEAT 4850 4872 LRR 216.  
 FT REPEAT 4872 4894 LRR 217.  
 FT REPEAT 4894 4916 LRR 218.  
 FT REPEAT 4916 4938 LRR 219.  
 FT REPEAT 4938 4960 LRR 220.  
 FT REPEAT 4960 4982 LRR 221.  
 FT REPEAT 4982 5004 LRR 222.  
 FT REPEAT 5004 5026 LRR 223.  
 FT REPEAT 5026 5048 LRR 224.  
 FT REPEAT 5048 5070 LRR 225.  
 FT REPEAT 5070 5092 LRR 226.  
 FT REPEAT 5092 5114 LRR 227.  
 FT REPEAT 5114 5136 LRR 228.  
 FT REPEAT 5136 5158 LRR 229.  
 FT REPEAT 5158 5180 LRR 230.  
 FT REPEAT 5180 5202 LRR 231.  
 FT REPEAT 5202 5224 LRR 232.  
 FT REPEAT 5224 5246 LRR 233.  
 FT REPEAT 5246 5268 LRR 234.  
 FT REPEAT 5268 5290 LRR 235.  
 FT REPEAT 5290 5312 LRR 236.  
 FT REPEAT 5312 5334 LRR 237.  
 FT REPEAT 5334 5356 LRR 238.  
 FT REPEAT 5356 5378 LRR 239.  
 FT REPEAT 5378 5400 LRR 240.  
 FT REPEAT 5400 5422 LRR 241.  
 FT REPEAT 5422 5444 LRR 242.  
 FT REPEAT 5444 5466 LRR 243.  
 FT REPEAT 5466 5488 LRR 244.  
 FT REPEAT 5488 5510 LRR 245.  
 FT REPEAT 5510 5532 LRR 246.  
 FT REPEAT 5532 5554 LRR 247.  
 FT REPEAT 5554 5576 LRR 248.  
 FT REPEAT 5576 5598 LRR 249.  
 FT REPEAT 5598 5620 LRR 250.  
 FT REPEAT 5620 5642 LRR 251.  
 FT REPEAT 5642 5664 LRR 252.  
 FT REPEAT 5664 5686 LRR 253.  
 FT REPEAT 5686 5708 LRR 254.  
 FT REPEAT 5708 5730 LRR 255.  
 FT REPEAT 5730 5752 LRR 256.  
 FT REPEAT 5752 5774 LRR 257.  
 FT REPEAT 5774 5796 LRR 258.  
 FT REPEAT 5796 5818 LRR 259.  
 FT REPEAT 5818 5840 LRR 260.  
 FT REPEAT 5840 5862 LRR 261.  
 FT REPEAT 5862 5884 LRR 262.  
 FT REPEAT 5884 5906 LRR 263.  
 FT REPEAT 5906 5928 LRR 264.  
 FT REPEAT 5928 5950 LRR 265.  
 FT REPEAT 5950 5972 LRR 266.  
 FT REPEAT 5972 5994 LRR 267.  
 FT REPEAT 5994 6016 LRR 268.  
 FT REPEAT 6016 6038 LRR 269.  
 FT REPEAT 6038 6060 LRR 270.  
 FT REPEAT 6060 6082 LRR 271.  
 FT REPEAT 6082 6104 LRR 272.  
 FT REPEAT 6104 6126 LRR 273.  
 FT REPEAT 6126 6148 LRR 274.  
 FT REPEAT 6148 6170 LRR 275.  
 FT REPEAT 6170 6192 LRR 276.  
 FT REPEAT 6192 6214 LRR 277.  
 FT REPEAT 6214 6236 LRR 278.  
 FT REPEAT 6236 6258 LRR 279.  
 FT REPEAT 6258 6280 LRR 280.  
 FT REPEAT 6280 6302 LRR 281.  
 FT REPEAT 6302 6324 LRR 282.  
 FT REPEAT 6324 6346 LRR 283.  
 FT REPEAT 6346 6368 LRR 284.  
 FT REPEAT 6368 6390 LRR 285.  
 FT REPEAT 6390 6412 LRR 286.  
 FT REPEAT 6412 6434 LRR 287.  
 FT REPEAT 6434 6456 LRR 288.  
 FT REPEAT 6456 6478 LRR 289.  
 FT REPEAT 6478 6500 LRR 290.  
 FT REPEAT 6500 6522 LRR 291.  
 FT REPEAT 6522 6544 LRR 292.  
 FT REPEAT 6544 6566 LRR 293.  
 FT REPEAT 6566 6588 LRR 294.  
 FT REPEAT 6588 6610 LRR 295.  
 FT REPEAT 6610 6632 LRR 296.  
 FT REPEAT 6632 6654 LRR 297.  
 FT REPEAT 6654 6676 LRR 298.  
 FT REPEAT 6676 6698 LRR 299.  
 FT REPEAT 6698 6720 LRR 300.  
 FT REPEAT 6720 6742 LRR 301.  
 FT REPEAT 6742 6764 LRR 302.  
 FT REPEAT 6764 6786 LRR 303.  
 FT REPEAT 6786 6808 LRR 304.  
 FT REPEAT 6808 6830 LRR 305.  
 FT REPEAT 6830 6852 LRR 306.  
 FT REPEAT 6852 6874 LRR 307.  
 FT REPEAT 6874 6896 LRR 308.  
 FT REPEAT 6896 6918 LRR 309.  
 FT REPEAT 6918 6940 LRR 310.  
 FT REPEAT 6940 6962 LRR 311.  
 FT REPEAT 6962 6984 LRR 312.  
 FT REPEAT 6984 7006 LRR 313.  
 FT REPEAT 7006 7028 LRR 314.  
 FT REPEAT 7028 7050 LRR 315.  
 FT REPEAT 7050 7072 LRR 316.  
 FT REPEAT 7072 7094 LRR 317.  
 FT REPEAT 7094 7116 LRR 318.  
 FT REPEAT 7116 7138 LRR 319.  
 FT REPEAT 7138 7160 LRR 320.  
 FT REPEAT 7160 7182 LRR 321.  
 FT REPEAT 7182 7204 LRR 322.  
 FT REPEAT 7204 7226 LRR 323.  
 FT REPEAT 7226 7248 LRR 324.  
 FT REPEAT 7248 7270 LRR 325.  
 FT REPEAT 7270 7292 LRR 326.  
 FT REPEAT 7292 7314 LRR 327.  
 FT REPEAT 7314 7336 LRR 328.  
 FT REPEAT 7336 7358 LRR 329.  
 FT REPEAT 7358 7380 LRR 330.  
 FT REPEAT 7380 7402 LRR 331.  
 FT REPEAT 7402 7424 LRR 332.  
 FT REPEAT 7424 7446 LRR 333.  
 FT REPEAT 7446 7468 LRR 334.  
 FT REPEAT 7468 7490 LRR 335.  
 FT REPEAT 7490 7512 LRR 336.  
 FT REPEAT 7512 7534 LRR 337.  
 FT REPEAT 7534 7556 LRR 338.  
 FT REPEAT 7556 7578 LRR 339.  
 FT REPEAT 7578 7600 LRR 340.  
 FT REPEAT 7600 7622 LRR 341.  
 FT REPEAT 7622 7644 LRR 342.  
 FT REPEAT 7644 7666 LRR 343.  
 FT REPEAT 7666 7688 LRR 344.  
 FT REPEAT 7688 7710 LRR 345.  
 FT REPEAT 7710 7732 LRR 346.  
 FT REPEAT 7732 7754 LRR 347.  
 FT REPEAT 7754 7776 LRR 348.  
 FT REPEAT 7776 7798 LRR 349.  
 FT REPEAT 7798 7820 LRR 350.  
 FT REPEAT 7820 7842 LRR 351.  
 FT REPEAT 7842 7864 LRR 352.  
 FT REPEAT 7864 7886 LRR 353.  
 FT REPEAT 7886 7908 LRR 354.  
 FT REPEAT 7908 7930 LRR 355.  
 FT REPEAT 7930 7952 LRR 356.  
 FT REPEAT 7952 7974 LRR 357.  
 FT REPEAT 7974 7996 LRR 358.  
 FT REPEAT 7996 8018 LRR 359.  
 FT REPEAT 8018 8040 LRR 360.  
 FT REPEAT 8040 8062 LRR 361.  
 FT REPEAT 8062 8084 LRR 362.  
 FT REPEAT 8084 8106 LRR 363.  
 FT REPEAT 8106 8128 LRR 364.  
 FT REPEAT 8128 8150 LRR 365.  
 FT REPEAT 8150 8172 LRR 366.  
 FT REPEAT 8172 8194 LRR 367.  
 FT REPEAT 8194 8216 LRR 368.  
 FT REPEAT 8216 8238 LRR 369.  
 FT REPEAT 8238 8260 LRR 370.  
 FT REPEAT 8260 8282 LRR 371.  
 FT REPEAT 8282 8304 LRR 372.  
 FT REPEAT 8304 8326 LRR 373.  
 FT REPEAT 8326 8348 LRR 374.  
 FT REPEAT 8348 8370 LRR 375.  
 FT REPEAT 8370 8392 LRR 376.  
 FT REPEAT 8392 8414 LRR 377.  
 FT REPEAT 8414 8436 LRR 378.  
 FT REPEAT 8436 8458 LRR 379.  
 FT REPEAT 8458 8480 LRR 380.  
 FT REPEAT 8480 8502 LRR 381.  
 FT REPEAT 8502 8524 LRR 382.  
 FT REPEAT 8524 8546 LRR 383.  
 FT REPEAT 8546 8568 LRR 384.  
 FT REPEAT 8568 8590 LRR 385.  
 FT REPEAT 8590 8612 LRR 386.  
 FT REPEAT 8612 8634 LRR 387.  
 FT REPEAT 8634 8656 LRR 388.  
 FT REPEAT 8656 8678 LRR 389.  
 FT REPEAT 8678 8700 LRR 390.  
 FT REPEAT 8700 8722 LRR 391.  
 FT REPEAT 8722 8744 LRR 392.  
 FT REPEAT 8744 8766 LRR 393.  
 FT REPEAT 8766 8788 LRR 394.  
 FT REPEAT 8788 8810 LRR 395.  
 FT REPEAT 8810 8832 LRR 396.  
 FT REPEAT 8832 8854 LRR 397.  
 FT REPEAT 8854 8876 LRR 398.  
 FT REPEAT 8876 8898 LRR 399.  
 FT REPEAT 8898 8920 LRR 400.  
 FT REPEAT 8920 8942 LRR 401.  
 FT REPEAT 8942 8964 LRR 402.  
 FT REPEAT 8964 8986 LRR 403.  
 FT REPEAT 8986 9008 LRR 404.  
 FT REPEAT 9008 9030 LRR 405.  
 FT REPEAT 9030 9052 LRR 406.  
 FT REPEAT 9052 9074 LRR 407.  
 FT REPEAT 9074 9096 LRR 408.  
 FT REPEAT 9096 9118 LRR 409.  
 FT REPEAT 9118 9140 LRR 410.  
 FT REPEAT 9140 9162 LRR 411.  
 FT REPEAT 9162 9184 LRR 412.  
 FT REPEAT 9184 9206 LRR 413.  
 FT REPEAT 9206 9228 LRR 414.  
 FT REPEAT 9228 9250 LRR 415.  
 FT REPEAT 9250 9272 LRR 416.  
 FT REPEAT 9272 9294 LRR 417.  
 FT REPEAT 9294 9316 LRR 418.  
 FT REPEAT 9316 9338 LRR 419.  
 FT REPEAT 9338 9360 LRR 420.  
 FT REPEAT 9360 9382 LRR 421.  
 FT REPEAT 9382 9404 LRR 422.  
 FT REPEAT 9404 9426 LRR 423.  
 FT REPEAT 9426 9448 LRR 424.  
 FT REPEAT 9448 9470 LRR 425.  
 FT REPEAT 9470 9492 LRR 426.  
 FT REPEAT 9492 9514 LRR 427.  
 FT REPEAT 9514 9536 LRR 428.  
 FT REPEAT 9536 9558 LRR 429.  
 FT REPEAT 9558 9580 LRR 430.  
 FT REPEAT 9580 9602 LRR 431.  
 FT REPEAT 9602 9624 LRR 432.  
 FT REPEAT 9624 9646 LRR 433.  
 FT REPEAT 9646 9668 LRR 434.  
 FT REPEAT 9668 9690 LRR 435.  
 FT REPEAT 9690 9712 LRR 436.  
 FT REPEAT 9712 9734 LRR 437.  
 FT REPEAT 9734 9756 LRR 438.  
 FT REPEAT 9756 9778 LRR 439.  
 FT REPEAT 9778 9800 LRR 440.  
 FT REPEAT 9800 9822 LRR 441.  
 FT REPEAT 9822 9844 LRR 442.  
 FT REPEAT 9844 9866 LRR 443.  
 FT REPEAT 9866 9888 LRR 444.  
 FT REPEAT 9888 9910 LRR 445.  
 FT REPEAT 9910 9932 LRR 446.  
 FT REPEAT 9932 9954 LRR 447.  
 FT REPEAT 9954 9976 LRR 448.  
 FT REPEAT 9976 9998 LRR 449.  
 FT REPEAT 10000 10022 LRR 450.  
 FT REPEAT 10022 10044 LRR 451.  
 FT REPEAT 10044 10066 LRR 452.  
 FT REPEAT 10066 10088 LRR 453.  
 FT REPEAT 10088 10110 LRR 454.  
 FT REPEAT 10110 10132 LRR 455.  
 FT REPEAT 10132 10154 LRR 456.  
 FT REPEAT 10154 10176 LRR 457.  
 FT REPEAT 10176 10198 LRR 458.  
 FT REPEAT 10198 10220 LRR 459.  
 FT REPEAT 10220 10242 LRR 460.  
 FT REPEAT 10242 10264 LRR 461.  
 FT REPEAT 10264 10286 LRR 462.  
 FT REPEAT 10286 10308 LRR 463.  
 FT REPEAT 10308 10330 LRR 464.  
 FT REPEAT 10330 10352 LRR 465.  
 FT REPEAT 10352 10374 LRR 466.  
 FT REPEAT 10374 10396 LRR 467.  
 FT REPEAT 10396 10418 LRR 468.  
 FT REPEAT 10418 10440 LRR 469.  
 FT REPEAT 10440 10462 LRR 470.  
 FT REPEAT 10462 10484 LRR 471.  
 FT REPEAT 10484 10506 LRR 472.  
 FT REPEAT 10506 10528 LRR 473.  
 FT REPEAT 10528 10550 LRR 474.  
 FT REPEAT 10550 10572 LRR 475.  
 FT REPEAT 10572 10594 LRR 476.  
 FT REPEAT

DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Probable serine/threonine-protein kinase pelle (EC 2.7.1.37).  
GN PLL OR CG5974.  
OS *Drosophila melanogaster* (fruit fly).  
OS *Eukaryota*; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_taxid=7227;  
RN [1]  
RN SEQUENCE FROM N.A., AND MUTAGENESIS.  
RP MEDLINE=93117834; PubMed=8440018;  
RX Shelton C.A., Masseman S.A.;  
RT "pelle encodes a protein kinase required to establish dorsoventral  
RT polarity in the *Drosophila* embryo."  
RL Cell 72:515-525 (1993).  
RN [2]  
RN SEQUENCE FROM N.A.  
RP STRAIN=Berkeley.  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amaratunga P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Bratton R.C., Rogers Y.-H.C., Blazer R.G., Champs M., Pfeiffer B.D.,  
RA Wan K.H., Doyle A.C., Baxter E.G., Helt J.G., Nelson C.R., Milos G.L.G.,  
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Baau A.A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Bemis P.V., Bertone J.P., Brockstein P., Brotier L.,  
RA Borkova D., Botchan M.R., Bouck Y., Brockstein P., Brothier A.,  
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dier S.M.,  
RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunko B.C., Dunn P.,  
RA Durkin K.J., Evangelista C.C., Ferraz S.C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., Mcherson D.,  
RA Merklow G., Mlshina N.V., Mobery C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy A., Nixon K., Nuskern D.R., Pacleb J.M.,  
RA Nelson D.R., Nelson K.A., Nixon K., Pollard J., Puri V., Reese M.G.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupeki M.P., Smith T.,  
RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,  
RA Strycharz R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach U.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye Y., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Zheng H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*."  
RL Science 287:2185-2195 (2000).  
RN [3]  
RN X-RAY CRYSTALLOGRAPHY (2.00 ANGSTROMS) OF 26-129 IN COMPLEX WITH TUBE  
RP MEDLINE=20055599; PubMed=10589682;  
RX Xiao T., Tomb P., Masseman S.A., Sprang S.R.;  
RA "Three-dimensional structure of a complex between the death domains  
RT of Pelle and Tube".  
RL Cell 99:545-555 (1999).  
RN [4]  
RN FUNCTION: REQUIRED FOR THE NUCLEAR IMPORT OF THE DORSAL PROTEIN  
CC WHICH ESTABLISHES DORSOVENTRAL POLARITY IN DROSOPHILA EMBRYOS.  
CC CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.  
CC -1- SUBUNIT: Interacts with tube through their respective N-terminal  
CC DEATH domains.  
CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT THE LIFE CYCLE WITH  
CC HIGHEST LEVELS IN 0-3 HOUR-OLD EMBRYOS AND ADULT FEMALES.  
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

CC PELLE SUBFAMILY.

CC -1- SIMILARITY: CONTAINS 1 DEATH DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; L08476; AAA28750.1; -

DR EMBL; AE003760; AAF56686.1; -

DR PIR; A45775; A45775

DR PDB; 1D2Z; 29-NOV-99.

DR PDBase; P8gn0010441; p11.

DR InterPro; IPR000488; Death.

DR InterPro; IPR002290; Ser\_thr\_kinase.

DR Pfam; PF00069; kinase; 1.

DR Pfam; PF00531; death; 1.

DR ProDom; PD000001; Euk\_kinase; 1.

DR SMART; SM00005; DEATH; 1.

DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.

DR PROSITE; PS00108; PROTEIN KINASE ST; 1.

DR PROSITE; PS00117; DEATH DOMAIN; 1.

DR TRANSFAM; Serine/threonine-protein kinase; ATP-binding;  
3D-structure.

FT DOMAIN 55 121 DEATH.

FT NP\_BIND 213 499 PROTEIN KINASE.

FT BINDING 219 227 ATP (BY SIMILARITY).

FT ACT\_SITE 240 240 ATP.

FT MUTAGEN 346 346

FT MUTAGEN 240 240

FT MUTAGEN 346 346

FT MUTAGEN 350 350

FT SEQUENCE 501 AA; 56160 MW; 4B292B40ACB81A8 CRC64;

Query Match 6.2%; Score 384.5; DB 1; Length 501;

Best Local Similarity 31.4%; Pred. No. 3.8e-15;

Matches 108; Conservative 75; Mismatches 132; Indels 29; Gaps 10;

QY 837 NSGRTNNTMKLTGVEKLS-----INLAEK---PLRLTPADLIQATNGPHND 887

DB 158 SSGVSNNNRSTTAAEIPSLISGNIHISYQRAASLEIDYALFNATDQSPDN 217

QY 888 LIGSGGFDVYKAILKDGSAVAIKLIHSGGQDREFMA-----EMETGKIKRNLP 941

DB 218 RLGGGGGVDYRGRKQKQ-LDVAIKVWYNSPNIDQKVELQSGYNELKYNIRHNDITLA 276

QY 942 LIGYCKAGDRLLVNEMVYKSLDELV-----QDPKGGVKKLSTRRKIAISGRGLAF 996

DB 277 LVYSYIKGGPCVLYVQIMKGSLEARLRAHKQNPPLA---LTMQQRSSISLGRGIYF 333

QY 997 LHNCSGPHIIHRMKSNNVLLENLEARYSDGMAFL-MSANDTHLSVSTAGTPGVVP 1055

DB 334 LHTARGTPLIHGDIKAPAMLLDQCLQPKIGDGLVBEGRKSLDAVENVKVFGRKYLPP 393

QY 1056 EYQSRGCTKGDVSYGVVLELLTGKRPDPSDFG--NNLVGMVQAHKALISVDFD 1113

DB 394 EERNFQSLSTGVDSYFGLVLEFVTRGVYDVRVENVETKQLDYAQQRNRNRELE 453

QY 1114 PELMKEDPALTELE-LQHLKAVAVACLDRAVRPRPTVQVMAMFK 1156

DB 454 KHL-AAPMKELDMCKALEGHLCTALDPODRPSMNAVLRKFE 496

RESULT 15

ID SH02 MOUSE STANDARD; PRT; 582 AA.

AC O88520;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE 15-JUN-2002 (Rel. 41, Last annotation update)

GN Leucine-rich repeat protein SHOC-2 (Ras-binding protein Sur-8).

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

ON NCBI TaxID=10090;

RP SEQUENCE FROM N.A.

RA MEDLINE=98337190; PubMed=9674433;

RT Stehrich D.S., Sun Q., Han W.;

RT "Sur-8, a conserved Ras-binding protein with leucine-rich repeats,

RT positively regulates Ras-mediated signaling in C. elegans.";

RL Cell 94:119-130(1998).

CC -1- SUBUNIT: SPECIFICALLY BINDS K-RAS AND N-RAS BUT NOT H-RAS IN

CC VIRO.

CC -1- SIMILARITY: CONTAINS 20 LEUCINE-RICH REPEATS (LRR).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC at the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AF068921; AAC40175.1; -

DR MGI; MGI:1921197; Shoc2.

DR InterPro; IPR001611; LRR.

DR InterPro; IPR003592; LRR out.

DR InterPro; IPR003591; LRR typ.

DR Pfam; PF00560; LRR; 17.

DR PRINTS; PRO0019; LEURICHRPT.

DR SMART; SM00370; LRR; 8.

DR SMART; SM00369; LRR typ; 5.

DR Repeat; Leucine-rich repeat.

FT REPEAT 99 122 LRR 1.

FT REPEAT 123 145 LRR 2.

FT REPEAT 146 168 LRR 3.

FT REPEAT 169 191 LRR 4.

FT REPEAT 193 214 LRR 5.

FT REPEAT 215 237 LRR 6.

FT REPEAT 239 260 LRR 7.

FT REPEAT 262 283 LRR 8.

FT REPEAT 284 306 LRR 9.

FT REPEAT 307 329 LRR 10.

FT REPEAT 331 353 LRR 11.

FT REPEAT 354 377 LRR 12.

FT REPEAT 379 400 LRR 13.

FT REPEAT 401 424 LRR 14.

FT REPEAT 425 447 LRR 15.

FT REPEAT 448 471 LRR 16.

FT REPEAT 473 493 LRR 17.

FT REPEAT 494 516 LRR 18.

FT REPEAT 518 540 LRR 19.

FT REPEAT 541 563 LRR 20.

SO SEQUENCE 582 AA; 64947 MW; 62C5C082B3CF5F12 CRC64;

Query Match 6.1%; Score 378.5; DB 1; Length 582;

Best Local Similarity 28.9%; Pred. No. 1.1e-14;

Matches 151; Conservative 80; Mismatches 188; Indels 103; Gaps 24;

QY 69 CRDQVTSIDISKPLVWGSFSAVSSSLSTGLSEPLFNSHNGSVSGFKGASLTSD 128

DB 97 CRENSMRDLDSKRSIH-----LPPSVKELTQLTLYLNSKIQSLPAVVGCLVNLMTLA 152

QY 129 LSRNSLGGVYTL-TSGSGSGGLKFLNVSNTL-DPPGVSGGLKLNLSLEVDLSANST 186

DB 153 LSENSL-----TSLPDSLDNKKLRMLDLRNKKLRIPSVV--YRLDLSLTLYLNFNIT 205

QY 187 GANYGVVLSD--GCGLKHLAISGNKISG-DVDVSRCNLEFLDVSSNNRSTGIPPLAD 243

Wed Mar 12 10:45:06 2003

us-09-823-394-2.rsp

Page 16

Db 206 -----TVEKDIKRLPYLSMISIRENKIKQLPRAIGELCNLITLVDVANHQ----- 249  
QY 244 CSALQHDIDISGNKLSGDFSAISTCTELKLIINISNQFVGPIPLPKSLQYLSLAENKE 303  
Db 250 ---LEHL-----PKEIGNCTQITNLDLQHNDLL----- 274  
QY 304 TGEIPDFISGACDTLTGIDLSGNHFYGAVPPFGSCSLFESLALSSNPSGELP---MDT 360  
Db 275 --DLPDTI-GNLSSINRLGLRYNRL-SAIPRSIAKCSALELLENNTIS-TLPESLSS 329  
QY 361 LKRGKLVLDLSF-----NEFS-----GELPESLTNLSASLITLIDLSN 400  
Db 330 LVKJNSLTARNCFOLYVGGPSOSTIYSLSMEHNRINKIPGIFERAKVLSKLNKDN 389  
QY 401 NFSGPIPLNLCNPRNTLOELYLQNGFTGKIPTLSNCELYSLHSFYLSGTI PSL 460  
Db 390 QUTS--LP-LDRGTWTSWELNLTATNQLT-KIPEVSGVSLLEVLIISNNILK-KLPHGL 444  
QY 461 GELSKLRDILKMLNMLEGEIIPQELMYVKTLETLIDFNDLTGEIIPSGLSNCTNIMWISLS 520  
445 GNLKRLKLELDLEENKLE-SLPNEIAYLKDOLQKLVITNNQLS-TLPBGIGHTNLTHTLGLG 502  
QY 521 NNRLTGEIIPWIGLENTLAILKLSNNSPESGNIPELGDORSU 562  
Db 503 ENLLT-HLPEIGTLENLELYLNDPNLHSLPFEIALCSKL 543

Search completed: March 10, 2003, 18:25:26  
Job time : 27.0143 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 18:11:50 ; Search time 34.2258 Seconds

(without alignment)  
3359.364 Million cell updates/sec

Title: US-09-823-394-2  
Perfect score: 6183

Sequence: 1 MKTFSSFFLSVTTLFFSF.....GSTTEMVDMSTKEVBEGL 1196

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

PIR 73:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6140	99.3	1196	2	T09356
2	2623.5	42.4	1166	2	F96598
3	2509.5	40.6	1143	2	B84431
4	1608.5	26.0	1192	2	T84899
5	1419	23.0	1232	2	T05322
6	1380	22.3	1079	2	C96772
7	1370	22.2	1108	2	D84434
8	1367	22.1	1133	2	B86308
9	1364	22.1	1124	2	B84742
10	1357.5	22.0	1003	2	T05898
11	1340.5	21.8	1093	2	G96746
12	1340.5	21.7	1134	2	T04587
13	1335	21.6	1104	2	B86465
14	1313	21.2	1013	2	T10659
15	1311	21.2	999	1	S27756
16	1310	21.0	1002	2	T46033
17	1296.5	21.0	1109	2	T18536
18	1293.5	20.9	992	2	T05335
19	1279.5	20.7	981	2	T50851
20	1273	20.6	1123	2	D96756
21	1257.5	20.3	996	2	F86410
22	1256.5	20.3	987	2	T50850
23	1230	19.9	1088	2	E86312
24	1237.5	19.9	976	2	B84659
25	1224.5	19.8	978	2	E96787
26	1222	19.8	932	2	T48489
27	1194.5	19.3	1027	2	B85089
28	1186.5	19.2	1120	2	B86479
29	1179	19.1	1029	2	T00712

Query Match	Best Local Similarity	99.3%	Score 6140	DB 2	Length 1196	Matches 1189	Conservative 0	Mismatches 7	Indels 0	Gaps 0
QY	1									
DB	1									
QY	61									
DB	61									
QY	121									
DB	121									
QY	181									
DB	181									
QY	241									
DB	241									
QY	301									
DB	301									
QY	361									
DB	361									

## ALIGNMENTS

RESULT 1	T09356	brassinosteroid-insensitive protein BR1 - Arabidopsis thaliana
N/Alternate names:	protein F23K16.30	
C/Species:	Arabidopsis thaliana (mouse-ear cress)	
C/Date:	11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 24-Nov-1999	
C/Accession:	T09356	
R/By:	Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.	
Submitted to the Protein Sequence Database, June 1999		
A/Reference number:	Z16652	
A/Accession:	T09356	
A/Molecule type:	DNA	
A/Residues:	1-1196 <BEV>	
A/Cross-references:	EMBL:AL078620; GSPDB:GN0062; ATSP:F23K16.30	
A/Experimental source:	cultivar Columbia; BAC clone F23K16	
C/Genetics:		
A/Map position:	4	
C/Suprafamily:	protein kinase Xa21; leucine-rich alpha-2-glycoprotein repeat homology; p	

QY 421 LYLONNGFTGKIPTLLSNCSSEVLSLHSPNYLSGTIPSSLSGLSKLRDLKMLNMLEGEI 480  
 Db 421 LYLONNGFTGKIPTLLSNCSSEVLSLHSPNYLSGTIPSSLSGLSKLRDLKMLNMLEGEI 480  
 QY 481 POELMYVKTLETLIIDFNDLTGEIPBGLSNTCNLAWISLSNNRLTGEIPKMTGRLENLAI 540  
 Db 481 POELMYVKTLETLIIDFNDLTGEIPBGLSNTCNLAWISLSNNRLTGEIPKMTGRLENLAI 540  
 QY 541 LKLSNNSFSGNIPDELGDRCRLTMDLNTNLFNQGTIPAMFQSGKILANFIAGKRYVYI 600  
 Db 541 LKLSNNSFSGNIPDELGDRCRLTMDLNTNLFNQGTIPAMFQSGKILANFIAGKRYVYI 600  
 QY 601 KNDGMEKECHGAGNILEFQGISSEQLNRLSTRNPNCTISRYVGHGHTSPPTDNNGSMFLD 660  
 Db 601 KNDGMEKECHGAGNILEFQGISSEQLNRLSTRNPNCTISRYVGHGHTSPPTDNNGSMFLD 660  
 QY 661 MSYNMLSGYIPKEISMPYLFILNLGHNDISGSIPEVEDLGLNLTLDLSSNKLDRIRPQ 720  
 Db 661 MSYNMLSGYIPKEISMPYLFILNLGHNDISGSIPEVEDLGLNLTLDLSSNKLDRIRPQ 720  
 QY 721 AMSALTMLTETIDLSNNNLSGPIPEMGOFETPPAKFLANPGLCGYLPKCDPSNADGVYH 780  
 Db 721 AMSALTMLTETIDLSNNNLSGPIPEMGOFETPPAKFLANPGLCGYLPKCDPSNADGVYH 780  
 QY 781 HOSHGRRPASLAGSVAMGLIFSFVCIIGLILVGRMKRRKKEALEMYAAGHNSGD 840  
 Db 781 HOSHGRRPASLAGSVAMGLIFSFVCIIGLILVGRMKRRKKEALEMYAAGHNSGD 840  
 QY 841 RTANNTWKLTGYVEALSTILAAFEKRLKLTADLLQATNGFHNDSLISGSGFGVYVA 900  
 Db 841 RTANNTWKLTGYVEALSTILAAFEKRLKLTADLLQATNGFHNDSLISGSGFGVYVA 900  
 QY 901 ILKXGSAVAIKGLIHVSGQDREFMAEMETIGIKHNLVPLLGYCKVGBERLLVNEWMK 960  
 Db 901 ILKXGSAVAIKGLIHVSGQDREFMAEMETIGIKHNLVPLLGYCKVGBERLLVNEWMK 960  
 QY 961 YGSLIEDVLQDPKKGAVKIKSTRKXIAISARGLAFLHNCSPHIHRDMKSSNVLLDEN 1020  
 Db 961 YGSLIEDVLQDPKKGAVKIKSTRKXIAISARGLAFLHNCSPHIHRDMKSSNVLLDEN 1020  
 QY 1021 LEAVSDPDMARLMSAMDTLHVSSTLACTPGYVPEYQSFRCSTKGDVYSYGVLLLEL 1080  
 Db 1021 LEAVSDPDMARLMSAMDTLHVSSTLACTPGYVPEYQSFRCSTKGDVYSYGVLLLEL 1080  
 QY 1081 TGKRPDTPDPEGDNNLVGMVYQHAQLRISDVPELMKEDPALISELQHLKVAACLD 1140  
 Db 1081 TGKRPDTPDPEGDNNLVGMVYQHAQLRISDVPELMKEDPALISELQHLKVAACLD 1140  
 QY 1141 RAMRPTVQVWAMPEKEIOAGSGIDOSTIRSIEDGFTIEMVDMISKEYPEGL 1196  
 Db 1141 RAMRPTVQVWAMPEKEIOAGSGIDOSTIRSIEDGFTIEMVDMISKEYPEGL 1196  
 RESULT 2  
 F96598  
 Protein F20N2.4 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: F96598  
 R:Phenology: A.; Ecker, J.R.; Palm, C.J.; Federici, N.A.; Kaul, S.; White, O.; Alonso,  
 R.; Chinn, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 ansen, N.F.; Hughes, B.; Huitrat, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marziani,  
 R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,  
 Ker, M.; Wu, D.; Xu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; M01D:21016719; PMID:11130712  
 A:Accession: F96598  
 A:Status: preliminary  
 A:Molecule type: DNA

A:Residues: 1-1166 <STO>  
 A:Cross-references: GB:AB005173; NID:q878502; PIDN:AAE79510.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F20N2.4  
 A:Map position: 1  
 Query Match 42.4%; Score 2623.5; DB 2; Length 1166;  
 Best Local Similarity 48.5%; Pred. No. 5.4e-112;  
 Matches 573; Conservative 190; Mismatches 341; Indels 77; Gaps 29;  
 QY 33 YREIHOLISFK--DVLDP-KNLLPDM--SSNKPCTFDEVTRGD-KYTSIDLSKPLNV 86  
 Db 32 FNEYALLAFKQSVKSDPNNVGNWKYBSGRSCSWRVGSCSDDERIVGLDRNSGLTG 91  
 QY 87 GFSAVSSLLSLTGLBSELEFNSHIN--GSVSGFKCSASTISDLSRNSLSGPTVLTSL 144  
 Db 92 TNLNV--NLTRLPYLOLTYLGNTYFSSGSDSSDC--YLOVTLDSNLSIDSVMDVYF 147  
 QY 145 GSCSGLEFLVNSVNTLDFPKVDS--GGIKNSLEVLDLSANSISGANVGVWVSDGGELK 203  
 Db 148 SKCSNLYSVNINNNKL--VGKGFAPSLQLTTPVDLSYNIISD-KIPESFISDFPASK 204  
 QY 204 HLAISGKISGDVDVSRVNLFLDVSSNNFSTGIPPLGDCSALQHLDISGKLSGD-FS 262  
 Db 205 YLDLTHNNLSGD-----FSDLS-----FGICMLTFPSLSQNNLSGDKFP 244  
 QY 263 RAISTCELEKLTINSNOFVGPFP-----LPLKSLQYLSAENKFTGEIPDFLSGACTL 318  
 Db 245 ITLPNCKFLETINISRNNAKIPNGEYWGSPQWLKQSLAHNRSLSGEIPPELSLCKTL 304  
 QY 319 TGLDSGNHFYGAVPFPGSCSLLESIALSSNNFSGELPMDTLLKMGGLVLDSENEFS 378  
 Db 305 VTLDSGNTFSGELPFGQACVWLQNLGNNTYLSGDFLNTVSKITGITLYLYVAINNIS 364  
 QY 379 GELPESLTNLASLTLDLSSNNFSGPIPLPVLCO--NPENTLOEYLQNGFTKIPPTLS 437  
 Db 365 GSVPISTNCS--NLKVLDSVSGFTGNVPSGFCLOSPPVLEKLLINNTYLSGVVMEIG 423  
 QY 438 NCSEVLSLHSPNYLSGTIPSSLSGLSKLRDLKMLNMLEGEIPQELMYK--TLETLIL 495  
 Db 424 KCKSLKTLIDLSNELLTGP1PKIIMWLPMLSLVMANNLTGP1P-EGCVVAGNMLETLIL 482  
 QY 496 DFNLDLTGEIPSGLSNCTNLMTSLSNNRLTGEIPKMTGRLENLAIILKLSNNSFSGNIPDE 555  
 Db 483 NNMLLTGSIPIESISRCTMWTWISLSNNRLTGEIPSGNLSKLTALIQGNNLSGNVPRQ 542  
 QY 556 LQDGRSLTMDLNTNLFNQGTIPAMFQSGKILANFIAGKRYVYIKNDGMEKECHGAGNL 615  
 Db 543 LGNCKSLTMDLNSNNLTGDLPGELASQAGLVMPGSVGKQPAFVRNEG-GTDCRAGAGL 601  
 QY 616 LRFQGISBQLNRLSTRNPNCTISRYVGHGHTSPPTDNNGSMFLDMSYNMLSGYIPKEIG 675  
 Db 602 VEPFGIRARLERLPMVHSCPAT-RIVSGMTWYTFPSANGSMYFIDISNAVSGFTPPGIG 660  
 QY 676 SMPYELTNLGHNDISGSIPEVEDLGLNLTLDLSSNKLDRIRPQAMGALTMLTETIDLSN 735  
 Db 661 NMGYLOVNLGNHRTLTGTPDSFGKALGVLDLSHNNLTGTLPSGLSGLSLDLDVSN 720  
 QY 736 NNTSGPIPEMGOFETPPAKFLANPGLCGYLPKCDPSNADGVYAHQSHGRRPASLAGS 795  
 Db 721 NMLTGP1PFGQGLTTPPVSRVYANNSGLCGVLPKPC--GSAPRPRTSRTHARK-QTVATA 777  
 QY 796 VAMGLIFSFVCIIGLILVGRMKRRKKEALEMYAAGHNSGDPTANTMKTGLGYE 855  
 Db 778 VIAGIAPSFMCFLMWALRYVRK--VQKQKQKKEKIEELPISG-----SCWKLSSVYE 831  
 QY 856 ALSTNLAAFEKRLKLTADLLQATNGFHNDSLISGSGFGVYKALKXGSAVAIKKLIIH 915  
 Db 832 PLSTNVATFEKRLKLTFAHLNLTNGFSAFTMVGSGGGEVYKQALRQGSVAIKKLIR 891  
 QY 916 VSGQDREFMAEMETIGIKHNLVPLLGYCKVGBERLLVNEWMKGSLEVDYQD--PKK 973  
 Db 892 ITGGDREFMAEMETIGIKHNLVPLLGYCKVGBERLLVNEWMKGSLETVLHEKSSK 951

QY 974 GGVKLUKSTRKKAIGARGIAFLHNHNCSPHIIHRDKSSNVLLDEBZARVDFGMARL 1033  
 Db 952 GGIIYMAARAKKIIAGARGLAFIHHSCIPHIIHRDKSSNVLLDEBZARVDFGMARL 1011  
 QY 1034 MSADTHLSVTLTGTGYVPEYVYOSFRCSYGDVYVYVLLTLLTKRPDPSDFG- 1092  
 Db 1012 VSADTHLSVTLTGTGYVPEYVYOSFRCSYGDVYVYVLLTLLTKRPDPSDFG- 1071  
 QY 1093 DNNLVGVKQ-HAKLRISDFVDEPELMKEDPALEIELLQHKVAACLDPRAMRPMTQV 1151  
 Db 1072 DNNLVGVAKQLYREKRGAEILDELY-TDKSGVELEFHYIKAKISQCLDRPRFRPTMQL 1130  
 QY 1152 MAMFKEIQASGIDOSTISIEDGCFSTIEMVDMKICEVP 1192  
 Db 1131 MAMFKEIMADTEDE-----SLDEPFLKERTP 1156

## RESULT 3

B84431  
 C:Accession: B84431  
 C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C/Accession: B84431  
 R./Lin, X./Kaul, S./Rounstey, S.D./Shea, T.P./Benito, M.I./Town, C.D./Fujii, C.Y./  
 M./Koo, H./Moffat, K.S./Cronin, L.A./Shen, M./Vanaken, S.E./Umayam, L./Tallon, L./  
 Neuser, D./Nieman, W.C./White, O./Eisen, J.A./Salzberg, S.L./Fraser, C.M./Venter, J.  
 Nature 402, 761-768, 1999  
 A/Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
 A/Reference number: A84420; MUID:20083487; PMID:10617197  
 A/Accession: B84431  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-1143 <SNO>  
 A/Cross-references: GB:AE002093; NID:g4406778; PIDN:AMD2008.1; GSPDB:GN00139  
 C/Genetics:  
 A/Genes: At2g01950  
 A/Map position: 2

Query Match 40.6%; Score 2509.5; DB 2; Length 1143;  
 Best Local Similarity 46.2%; Pred. No. 8.1e-107;  
 Matches 556; Conservative 183; Mismatches 353; Indels 111; Gaps 25;

QY 9 LSVTTLFFPSFSLSFQASQSLYREIHQIISFQVLPD-KULLPDMSNKRPCTFDG 66  
 Db 14 IQISFIFLTLHSGS-SSDDOSLKTDSLISFETKMLQDDPNILSNMSPKSPCCQFSG 72  
 QY 67 VTCRDDKVTSLDLSKPLN-VGSAVSS-SLSTLTGLESLFLSNHSHNGSVGFKC 120  
 Db 73 VTCLOGRVTETIMSSGSGIIVSFNAFTSLDSLAVLKLS-ENFVFN-----STSLILL 125  
 QY 121 SASLTSLSLDSRNSLGGPVTTTSLGSCGSLKFLVSSNTL-DPFGKVGGLKANS-L 175  
 Db 126 PLTLVHLEHSSSGLG-----TLPENFSSKYSNLTSLTSSYNFETKLPNDLFLSKKL 179  
 QY 176 EYLDISANSISGANYGVVSLDGGELKHLAISGNKISGVDVSRCVNLEFLVSSNNS 235  
 Db 180 QTLDSYNNITG-----PISGLTLP----- 199  
 QY 236 TGPIPLGDCSALQHLIDISGNKLSGDFSAISTCTELKLNISNQVGPPIPL-PLKSL 293  
 Db 200 -----LSSCVMTYLDREGNSISYIDSLDNLCTNKLMSYNFPGQPKSFGELKLL 254  
 QY 294 QYLSLAENKFFGEIPDFLSGACDTLTGLDSGNHFGVAVPPFGSCSLSESLSSNNS 353  
 Db 255 QSLDLSHRLTGWIPPEIGDTCRSLQNLKRSYNNFTVPESSSCSWLQSLDLSNNTS 314  
 QY 354 GELPMDTLTKRGKLVLDLGFNEBSGELPESLTNLSA-SLTLTDLSSNFGSPLIPNL 411  
 Db 315 GPPFTVITIRSGSQIITLISNNLTSGDFP--TISACKSLRIADPSSNRFGSVIIPDLC 371  
 QY 412 QNPKNLTQELVQNNNGFTGKIPPLTSLNCSSELVSLHSFNLSTISSTSSGSLKDLKL 471

Db 372 PGAA-SLEELRLDNLVTGERIPPAISQCELELTIDLSLNYLNGTIPPEIGNLOKLEOFA 430  
 QY 472 WLNLEBEPIDELMYVKTETLLIDPNDLTGEIPSGLSCTNMIANISNNRLTGEIRKM 531  
 Db 431 WYNNIAGEIPPEIGKLONLKDLILANNQLTGEIPPEFSCNIEBVSFTSNRLTEVRK 490  
 QY 532 IGRLENALIKLSNNSPSCNIPDELDCSLWLDLNTLNFNGTIPAAFRQSG-KLAN 590  
 Db 491 FGLISRLAVQLGNNNFTBEIPPEIGKCTTLVWLDLNTLHLEIPPEIGPQSGKLSG 550  
 QY 591 FIAGKRVYIKNDGMKECHGAGNLEFQIGRSBQNLSTBNPNTISRYGGHTSTPE 650  
 Db 551 LLSGNMAFVAVNG--NSCKGVGLVEBSGIRPERLQIPSLKSCDT-RMYSGTISLF 607  
 QY 651 DNNGSMFIDMSYNNLSGYIPKEISMPYLLFNLGNDISGSIPEVDGRGLNIIDLS 710  
 Db 608 TRYQITELDSLSTQNLGKIPPEIGEMIALQVLELSHQNGEIPFTIGQKNGVDPAS 667  
 QY 711 SNKDRIPQAMSLTMTLEIDLSNNLSGPIPEMGQETPPAKFLANPGLCGYPLRC 770  
 Db 668 DNLQOQIPESFSNLSFLVQIDLSNNELTGPDPKGQSLTPAQYANNPGLCGVPLPEC 727  
 QY 771 DPNAD--GYAHQRS-HGRPPASLAGVAMGLFSPVCIIFGLILVGRMRKRRKKEA 826  
 Db 728 KNGNNQLPAGTEEGRAKGRPAASMANSYLVGLVISAASCIILVAIAVARARRDAD 787  
 QY 827 ELEMYAEGHNGSDTANNYKLTGYKALSLINLAPEKPLRLTPADLLQATNGFHD 886  
 Db 788 AKMLHSLQAVNSA-----TWKIEKEKEPISINVAIFQRLKFSQILEATNGSSAA 841  
 QY 887 SLISGSGFDGVYKALIKDGSAAVAKKLIHVSGODREFPMEMETIGIKIRNVPLGLYC 946  
 Db 842 SMIGHGGEVGVKATLKDGSSVALKILIRLSCQDRFEMEMETLGKIKRNVPLGLYC 901  
 QY 947 KYGDELLVNEVMKYGSLEEDVLODPKKGVK--LKLSTRKKAIGARGIAFLHNHNCSPH 1004  
 Db 902 KIGBERLLVYEMQYSLEEVLAGHGPFGKEKRIIGWERKIKAKGKICFLHNHCIPH 961  
 QY 1005 IHRDMKSSNVLLDENLEARYSDRGMARLMSAMDTLHLSVTLTGTGYVPEYVYOSFRCS 1064  
 Db 962 IHRDMKSSNVLLDQDEARVSDGMARLISALDTHLSVTLTGTGYVPEYVYOSFRCS 1021  
 QY 1065 TKGDVYSYGVVLELTLGKRETVDPDFGDNLLVGWVYQKHKL-RISVPPBELMKEDPL 1123  
 Db 1022 AKGDVYSYGVVLELTLGKRETVDPDFGDNLLVGWVYQKHKL-RISVPPBELMKEDPL 1081  
 QY 1124 EI-----ELLQHKVAACLDPRAMRPMTQVYAMFKEIQASGIDOSTI 1170  
 Db 1082 SLNEKEGEGGVIVKEMIRYLRLALRCVDDPPEKRPVMLQVAVSLRELR-GSENNSHSHS 1140  
 QY 1171 RSI 1173  
 Db 1141 NSL 1143

## RESULT 4

148499  
 receptor-like protein kinase-like protein - *Arabidopsis thaliana*  
 M/Alternate names: protein T28U14.220  
 C/Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C/Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
 C/Accession: T48499  
 R./Bevan, M./Murphy, G./Ridley, P./Hudson, S./Bancroft, I./Mewes, H.W./Rudd, S./Len  
 submitted to the Protein Sequence Database, April 2000  
 A/Reference number: Z24493  
 A/Accession: T48499  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-1192 <BEV>  
 A/Cross-references: EMBL:AL163652  
 A/Experimental source: cultivar Columbia; BAC clone T28U14  
 C/Genetics:  
 A/Map position: 5

A&gt;Note: T28J14.220

Query Match 26.0%; Score 1608.5; DB 2; Length 1192;

Best Local Similarity 35.2%; Pred. No. 8.5e-66; Mismatches 421; Indels 225; Gaps 36;

Matches 453; Conservative 177; Mismatches 421; Indels 225; Gaps 36;

11 VTTLPFFSFFSLSFQASPSQSLYREIHQISFQDVLDPKULLPDW--SSNNKPTCFDQVT 68  
 4 LTMFLFLFF--SFSSALVDLSSETTSLISFRRLENSLLSSNNVSSASASHCDWVGVT 61  
 69 CRDDKYSIDLSKPLNVGSAVSSLSITGLSEFLPNSHNGVS---GFCASAL 124  
 62 CLIGRVNSLSIPLSIR--GQIPKEISSLKNNRELCLAGNPGSKIPPEIMNK---HL 115  
 125 TSLDLSRNSLSGVVTLTSLGSCGKFLVNSNTLDPFGKV--SGGLKLNLSLEVLDSA 182  
 116 QTLDDLGSNSLTGLLPL--LSELPQLLYDLSN--HFGSLPSPFFISLPALSSLDVSN 171  
 183 NSISGANVGVWVLSDGGELKHLA--ISGNKISGDV----- 216  
 172 NSLSGE-----IPPEIGKLSNLSNLMGLNSFGQIPSEIGNISLKNPAAPSCFNGP 225  
 217 ---DVSRCVNLDFIDVSSNNFSTGI-----PFGDCSALQ 248  
 226 LPKXISIKLKLAKLDLSYNPKCSIPKSFGBLHNLSTLNVSAELGLIPPELGNCKSLK 285  
 249 HLDISGNKLSGDFRAISTCTELKLNLS--NQFVGPPIPL--PLKSIQYLSLAENKPT 304  
 286 SLMSFNSLSGPLLELS--EIPLPFSARNOGSLSPSMQKMWVLDLSLLANNRFS 342  
 305 GEIP-----DFLSGA-----CDTLGDLDSGNHHYGVVPPFGSCSL 341  
 343 GEIPHEIEDCPMLKHLASLNSLSSSIPRELCSGSLAIDLNSLISGTLIEVDFGCCS 402  
 342 LBSLALSSNNSGELPMOTLLKMGKLVLDLSPNBSGELPEST--TNL----- 388  
 403 LGEILLTNQNGISIPED-LMKL-PMALDDLDSNNFGEIPKLSMKSTNMEFTASYNRL 460  
 389 -----SASILLTLDSSNNFSGPILPNLCONPKNTLOELYLQNNNGPFGKIPPTLSN 438  
 461 EGYLPABIGNANSLKRLVLSNQLGGEIPREI--GKLTSLSVLNMANNMFOGKIPVELGD 518  
 439 CSELVSLHSPVLSGTSSLSLSISKLRLDKLMLMEGEIPQ-----ELMY 486  
 519 CTSLLTLDLGSNNLQGGIPDKITLALOQLCVLSTNNLSGSLPSKPSAYRHQIEMPLSF 578  
 487 VTTLETLIDFNDLTGEIPSGISNCTNINWISLNNRLTGEIPKWTIGLENLAILIKLSNN 546  
 579 LQHHGIPDLSTYRSLSGPIPELGEGLVVEISLNNHLSGEIPASLSRLTULTLIDLSGN 638  
 547 SFSGNIPDELGCRLIMDLNTNLFNCTIPPAAMFGKSGKIAANFIACKRYVYIKNDGMK 606  
 639 ALTGSIPEKMGNSLKLOGINLANNQINGHIP----- 669  
 607 KCHGAGNLEFGGRSEQNLRLSTENPCNITSRYGHTSPTFDNNGSMFELMSYML 666  
 670 -----ESFGLISLVKLNLTNKKLDEGVPSASLONKELTMDLSFNNL 712  
 667 SGYIPKEIGSMPLYFLINLGHNDISGIPDEVGDLRGNLIDLSNKLDRIPQASMLT 726  
 713 SGELSELSTMEKLVLYEONKFTGEIPSELGNLTQLELYLVSNLLSGELPTKICGP 772  
 727 MLTEILSNNNSLGGPIPEMGQFETPPPAFLANPGLCGPLRCDPSNADGVAHHORSHG 786  
 773 NIEFLINAKNNLRGEVPSDGVCDPSKALLSGKELCG--RVGSDC-----KIEG 821  
 787 RRPASLAGVAMGLFS---FVCI FGL--ILVGERMRK--RRKKEALEMYAEG--HG 836  
 822 TKLRASWAGIAGMLGFIITVVFPSLRAMWAKRYKQDDPERMESRLKGVQDNLTF 881  
 837 NSGDRANTNWKLTGVKEALSNLAAFEKPLKLTADILQATNGFNHDSLIGSGRSD 896  
 882 LSGSRS-----REPLSINIAMFEQPLKVLKVLGDIVETDHFSSKNITIGDGFET 930

QY 897 VYKAILKDSANVAIKTLINVSOGDREEMAEETIGKIKRNLVPLAGYCKVGBRELLVN 956  
 DB 931 VYKACLPGEKTVAVKLSSEAKTOGNREPMEMETLQKVGHPNLVSLGCSFSEELVY 990  
 QY 957 EWMKYSLEEDVLQDPKKGVKVLSTRKRIAGSARGALFELHNSCPHIIHRDKMSNVL 1016  
 DB 991 EYMWNSLDHMLNQGMLVLDMSKRLKIVAGARGALFLHGFIPHIHRDKASNTL 1050  
 QY 1017 LDENLEARVSDPGMARLMSMDTHLSVSTLACRPGVPEPYQSPFCSTKGDVSYGVVL 1076  
 DB 1051 LDGDFEPRVADFGIARLISACESHVS-TVLAGTFGTPPEYQASARATTKGDVSYGVVL 1109  
 QY 1077 LELLTGKRPDSDPGDN--NLGWYKQCHAKL-RISDVPPELMKEDPALTEILLQHL 1132  
 DB 1110 LELVTGKEPT-GDFPESBEGNLVGMALQKINGKQAVDVIDPLV--SVALKNSQURLIQ 1166  
 QY 1133 VAVACLDDRAWRRPTWQVYAMFKEI 1158  
 DB 1167 IAMCLAEETPAKRPNNMLDVALKALKEI 1192

## RESULT 5

T05322

hypochemical protein F18F4.240 - Arabidopsis thaliana

N.Alternate names: hypochemical protein F1C12.60

C.Species: Arabidopsis thaliana (mouse-ear cress)

C.Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 31-Jan-2000

C.Accession: T05322

C.Description: T05322

A.Reference number: 215408

A.Accession: T05322

A.Molecule type: DNA

A.Residues: 1-1232 &lt;BEV&gt;

A.Cross-references: EMBL:AL022224

A.Experimental source: cultivated Columbia; BAC clone F1C12

A.Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.

submitted to the Protein Sequence Database, February 1998

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A.Cross-references: EMBL:AL021637

A.Experimental source: cultivated Columbia; BAC clone F18F4

C.Description: T04898

A.Reference number: 215388

A.Accession: T04898

A.Molecule type: DNA

A.Residues: 1-305 &lt;BEV&gt;

A

```

Db 233 AELGRLENIETIANNSUTGE-----IPQLGMSQLOYLISLMAVLOQLIFKSIADL 286
Qy 222 VNLLEFLDVSSNNFSTGIP-FLGDCSALOHLISGNKLSGDSBRAI-STCELEKLINISSN 279
Db 287 GNLQTLDSANNLGEIPEEFNNNSQLDLVLANNHSGSLPKSICNNNTNLEOLVLSGT 346
Qy 280 QPVGIRPP-LPLKSLQYLISLAENKFTGEIPDFLSGACDTLTGLDLSGNHFGYAVPPFG 337
Db 347 QLSGELIPELSKQSLKQDLDSNNSLSGSIPEALFEIVE-LTDLYLNNTLEGLSPSIS 405
Qy 338 SCSLESLATSSNNFSGELPMDTLTKRGKLVLDLSNERSGELPESLTLNLSLTLDI 397
Db 406 NLTLNQMLVNLHNLLEGLPKR-ISAIRKLEVLPLYNRRSGELPOEIGNCT-SLMKIDM 463
Qy 338 SSNNFSGELIPLNLCQNPNTLOELYLQNNNGFTGKIPPTLSNCELSVLSHFSYLSGTIP 457
Db 464 FGNHFEGERIPPSI--GRUKELNLTALHONELVGLPRLSLGNCQNLNLDLADNOLSGSIP 521
Qy 458 SLSLSLKLRLDKLMLNMLBGEIPELMAVYKTELTLLDPNDLTG----- 502
Db 522 SSFGLKLEBOLMLYNNLSQNLGPDLSLRNLTNLSHNRNGTTHPLCGSSVLSYSD 581
Qy 503 -----ELPSGLSNCNTNLTWISLNNRLTGEIPKIGRLNLTATLKSNNSPGNTIPD 554
Db 582 VTNNGFEDEIPELGNQNDLRLGKNQUTGKIPTWLGKIRLSLDMSSNALGTGTP 641
Qy 555 ELGDCRLIWLDTNLTNCTIIPAMEFK--QSGKIANFJAGRYYIKNQDKKECHGA 612
Db 642 QLVLCRKLTHIDLNNTNPLSGRIPPWLGKLSQLGELK--LSSNQFV--BSLPELPLNC 694
Qy 613 GNLL-----EFGIRSEOLNLTSTRNPNITSRYGHTSPTPNNSSMFLDMSTYM 665
Db 695 TKLLVSLDGNLSNGSLPOEIGNLGALNLTNDKQFSSLPQAMKLSKLYELRLSRS 754
Qy 666 LSGYIPEIGSMYPL-FITNLGNDISGSIPEVGLRGLNLTLDLSNKLDRIGQANSA 724
Db 755 LTGEIPEIQLODLOSLDLSYNNFTGDIPTIGTSLTLETLDDLSHNOJGEVGSVGD 814
Qy 725 LTVLTLDSNNNLSGRIPBMOGFTFPRAKTLNPGCLCGYLPFCDBSNADGVAHORS 784
Db 815 MKSLGYLNVSPNNLGGKLRK--QFSRMPADSFGLGNTGCGSPLSRCN----- 859
Qy 785 HGRRLASLGSVAMGLFSEVFCIFLIVGRMRKRR--KKEALEMYAEGHNSGDR 841
Db 860 ---RVRTISALTAIGMLIVLAF-----FKORDFFK-----VHGSTAYT 899
Qy 842 TANNTWKLTGYKALSLINLAFEPKRLK-----LTFADLLQATNGFNDLSIGSGF 894
Db 900 SSSSSS-----QATHKPLFRNGASKSDIMMEDIMETNHLSEBPMISGGS 945
Qy 895 GVVYKAILKDSAVAIIKLIHVS--QGDREPAEMETIGKIKHNLVPLIGCYVADG-- 951
Db 946 GAVYKALENGTAVAKILMKODLMSNKSFSREYKTLGRIRHRLVLMGYSCKSEGL 1005
Qy 952 RLLVNEVWYGSLEVDLPK-----KGVKYLKLSRRKIIAGSARGLAFLHNSGPIIH 1007
Db 1006 NLLIYEYMGSGSIMWMLHEDKRVLEKKKLLDMWAKRLIANGLAQGYEYIHLHDVPIIVH 1065
Qy 1008 RDMKSNVLLDENLEARSDDPGMALMSA-MDTHLSVT-LAGTPGYVPEYVQSPFCST 1065
Db 1066 RDKISSNVLLDSNNMAHIGDGLKAVLTENDCTIWTDSWTWPAFCYGIABEYVYSKATE 1125
Qy 1066 KGDVYSYGVVLELLTGRPTDSDFGDNNTLVGVWKOAKLRIS--DVDDPELMKEDPA 1122
Db 1126 KSDVYSWGIWMEIYTGMPDTSVFGAEMDNVWETHLEVASARDKLDLPKLP 1185
Qy 1123 LEIELQLKVAVACLDLRAMRPTMVO 1150
Db 1186 EBDACQVLEIALQCTKTSPOERPSRQ 1213

```

RESULT 6

```

C96772
probable receptor protein kinase FIM20.4 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_rev: 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C96772
R:Theologian, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luos, U.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.
ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: C96772
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1079 <SNO>
A:Cross-references: GB:AB005173; NID:96539238; PIDN:AAF15908.1; GSPDB:GN00141
C:Gene: FIM20.4
A:Map position: 1
Query Match 22.3% Score 1380; DB 2; Length 1079;
Best Local Similarity 32.2%; Pred. No. 1.9e-55;
Matches 392; Conservative 204; Mismatches 406; Indels 214; Gaps 44;
Qy 26 ASPQSGLYREIHQLISFKVLPDKR-----LLPDKSSKNP--CTPQVTC--RDKYTS 76
Db 5 AVAGDLSDDREVLISLKSYLESRNPQKRGLYTEKMNQVVCQWPEITCTPPRSRYTG 64
Qy 77 IDLS-----SKPLNVSFVSSLSLTGLSLSLNSHNSVSGFKSASLTSLDSRN 132
Db 65 INLTSTISGPFKNSALT-----ELTYLDSRN 94
Qy 133 SLSPYTLTSLGSCSGKPLNVSSNTLD-----PPKYSGLKLNLSLEVDLSANISGA 188
Db 95 TIEGEIP--DLSRCHNLKHLNLSHNLLEGLSLPG-----LSNLEVDLSLRITGD 145
Qy 189 NVVGVNLSDCGCELKHLAISGNKISGVD--VSRVNLLEFLDVSSNNF-----TGI----- 238
Db 146 IGSFPLF--CNSLVANLSTNNFTGRIDIDFNGCRKLKYVDPSNRFSEWWTGREGLV 203
Qy 239 -----PFLGCSALOHLISGNKLSGVD--VSRVNLLEFLDVSSNNF-----TGI----- 282
Db 204 BPSVADNHLSGNISASMERGNT--LQWLDLSGNAFGGEFPGQVSNCONINVLNLMGNKFT 262
Qy 283 GPIPL--PLKSLQYLISLAENKFTGEIPDFLSGACDTLTGLDLSGNHFGYAVPPFGSGS 340
Db 263 GNIPAEIGSISLKGVIYGNTPSRDIPETLNLN--LVFLDLSRKKFGGDIOEIRGFT 321
Qy 341 LLESIALSSNNFSGELPMDTLTKRGKLVLDLSFNEPSGELPESLTLNLSLTLDI 400
Db 322 QVKYLVHANSYGVGINSNLTLPNLSRLDLYGNFNSGGLPREISQI----- 369
Qy 401 NFSGELIPLNLCQNPNTLOELYLQNNNGFTGKIPPTLSNCELSVLSHFSYLSGTIP 460
Db 370 -----QSLKFLILAIYNNFSGDIPQEGYNNPGLQALDLSFNKLTGSIPIAF 414
Qy 461 GSLKRLDKLMLNMLBGEIPELMAVYKTELTLLDPNDLTGIPSGCLNNTNLMWISIS 520
Db 415 GKLTSL--LM-----LMLANNLSGRIPEIGNCTSLMFWPA 450
Qy 521 NNRLTGEIPKIGRL--ENLAILKLSNNSPSGNIPDELGDCSL--IWLDTNLTNFGTII 578
Db 451 NNQSGRFPHELTRMGSNPSPTPEVVRQNKDKIAGSGCLAMKWIIPAEFPFR--FYVA 509
Qy 579 AMFKSGKIANFJAGRYYIKNQDKKECHGAGNLLEFQIRSEQLNRLSTRPCNIT 638
Db 510 ILTKGCSRLMDHV-----LKGYGLFPVC--SAGST-----VRTLKISAY-----QLS 551

```

QY 639 SKVGHSTPTEDNNGSMFMDMSYNNLSGYIPKEIGSMPLYLILNGLHNDISGIPDEV 698  
 Db 552 GNKFSGEIPASISQMDRSLTTLHGFNEFBGKLPPEIGQLPLAF-INLRNNNSGSEIPEI 610  
 QY 699 GDLRGNTLIDLSNNKLDGRIPQMSALTMLEIDLSNN-LSGPPEMGQETFPAPAF 757  
 Db 611 GNLKCLQNDLSFNNNSGNFPTSINDNLSEKFNISYNFISGALPTTGQVAFPKDSFL 670  
 QY 758 NNFGLCGYLPREDPSNAGYAHQSH---GRRPASLA---GSVAMGLFFGVCIFGLI 811  
 Db 671 GNP-LIRFP---SFPNGSNMTRKLSNOYLRKRPFTLLIIMISLALAF---IACIV 721  
 QY 812 LVREMRKRRKKEALEMYAEHGNSGDRJAN---NTWKLTGYKEALSINLAEPKL 868  
 Db 722 VSGIVLVVAKASREABDL-LDSKTRHDMTSSGSSSPM-LSGKIKVIRLDPKSTP--- 775  
 QY 869 RKLTFADLLQATNGFHNDSLISGSGFGDYKALKKQSGAVAKKLHVSQGGQREMAEM 928  
 Db 776 ---TYADILKATSNFSEHVRVGGYGTIVRGVLPQREVAVKLQREGTEAKEPRALM 832  
 QY 929 ETI-----GKIKRNLVPLLYGCKVDEBLVNEVMKYSLEBDLDPKKSGVKLSTR 983  
 Db 833 EVLSANAFGMAHPNVLRLGWCLDSEKILVHEMGGSLLEELID-----KTKLQWK 886  
 QY 984 RKIAIGS--ARGLAFLHNSCPHIIHRDMKSSVLLDENLEAVSDFGMARLMSANDTHL 1041  
 Db 887 KRIDLADVARGLVFLHECYPSIVHRDVKASVLLDKHGNARVDFGLARLNLVGDISHV 946  
 QY 1042 SVSTLACTPGVPPVPEYQSPFCSTKGDVSYGVVLLLELTKGKPTBSPDGDNNLVGMVK 1101  
 Db 947 S-TYIAGTIGVAAPEYQGTQATRRGDVSYGVVLMELALGRVADG---GEBELVEMAR 1002  
 QY 1102 QHAKLRISDVDPPELM---KEDPALTEILLQHLKVAAVLCDRAARPPVAVQVMAFEKI 1158  
 Db 1003 RVTMGNTAKSPITLSGTRKNGAE-QMTELLKIGVCTADHPQAPNMKEVLAMLVKI 1061  
 QY 1159 ---QAGSGIDSQSTI 1170  
 Db 1062 SGKAEFLNGLSQGYI 1077

RESULT 7  
 D84434 Probable receptor-like protein kinase [imported] - Arabidopsis thaliana  
 C.Species: Arabidopsis thaliana (mouse-ear cress)  
 C.Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C.Accession: D84434  
 P.O.O., X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 Doo, H.; McFar, K.S.; Cronin, L.A.; Shen, M.; VanKen, S.E.; Unayam, L.; Tallon, L.;  
 D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A.Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A.Reference number: AB4420; MUID:20083487; PMID:10617197  
 A.Accession: D84434  
 A.Status: preliminary  
 A.Molecule type: DNA  
 A.Residues: 1-1008 <STO>  
 A.Cross-References: GB:AE002093; NID:96598459; PIDN:AACT8507.2; GSPDB:GN00139  
 C.Genetics:  
 A:Gene: At2g02220  
 A:Map position: 2

Query Match 22.2%; Score 1370; DB 2; Length 1008;  
 Best Local Similarity 31.4%; Pred. No. 4.9e-55;  
 Matches 384; Conservative 181; Mismatches 359; Indels 300; Gaps 34;

QY 8 FLAVTTLFFSFSLSFOASPSQSLVR-ETIQLISFKDVLDPKULLPD-W---SSKNKNC 62  
 Db 12 FLTELCLFYS-----SESQTSRCHPHDEALRDFAHLEPKPDGINSSTDCG 63  
 QY 63 TPDGVTCRDKVTSIDLSKPLNVGSAVSSLSLTGLESLPSHSHNGSVSGKCA 122  
 Db 64 NMWGITC-----NSNMTGRV----- 78

QY 123 SLTSLDLSRNSLSGPVTLTLISGSCGLKFLNWSNTLDPKGVSGGLKLNLEVLIDJA 182  
 Db 79 --IRLEGNKKISGKLS--ESLS-----KLDIEIRVLNISR 109  
 QY 183 NSISGANVGVVLDSCGELKHLIASGNKISGDVDVSRVNLFEVDVSNNSFTGIPFLG 242  
 Db 110 NFI-----KDSIPLSIFNKNLQITLIDSSNDLSGGIPSI 144  
 QY 243 DCSALQHLDISGNKLSGDF-SRAISTCTELKLNLSNSQFVGPPIPLPKSLQYLSAEN 301  
 Db 145 NLPALQSFDSLSSNKNFSGSPSHICHNSTQIRVVKLAVENTFAG----- 186  
 QY 302 KFIQEIPLDPSGCDTLTGDLISGNHFGAVPPFSGCSLLESALSSNNFSGELPMDTL 361  
 Db 187 -----NFTSG-----FGCVLLEHLCIGMNDLTGNIPED-L 216  
 QY 362 LKMRGLKVLDSFNEFSGELPELITNLSALTLTLDSSNNFSGPI-----LPNLQNPK 415  
 Db 217 FHLKRLNLGIGENRLSGSLREIRNLS-SLVRLDVSWNLFSGEIPDPDELPLQ----- 270  
 QY 416 NTLQELVI-QNNGETGKIPTLNSCSELVSLHSFNYSGLTIPSSLSGLSKLDLKLMLN 474  
 Db 271 ---KFLGQTNFGTIGIPKSLANSPLNLNLNNSLSGRIMNCTAMALNSLDGTN 326  
 QY 475 MLEGEIPOELMYVKTLETLTLLDFNDLTGEIIPSGLSNCTNLNLSLN----- 522  
 Db 327 RENGRLPRLNLDCKRLKNVNLARNTFGQVPESEKNEFSLSYFSLNSLSANISSALGIL 386  
 QY 523 -----RLTGEIPMIGRELENLAIKL 543  
 Db 387 QHCKMLTTLVTLNFGHALPDSSLHFEKLVANCRILTSGMPRLSSNLDQLDLD 446  
 QY 544 SNNSPSGNIPBELGCRSLIMLNTLNFNGTIPAMFK-----OSGKIAPNFAKERYV 599  
 Db 447 SNMRLTGALPSYIGDFKALFYIDLSSNSFTGEIPKSLTKLESLSRNISVNEPDPPEFF 506  
 QY 600 IKNDGKKECHGAGNLBFGQIRSEQLNRLSTRNCPNTSVYGVGHSTPTEDNNGSMFL 659  
 Db 507 MKRN-----ESARALOYN-----OIFG--FPPF-----I 528  
 QY 660 DMSYNNLSGYIPKEIGSMPLYLILNGLHNDISGIPDEVGDLRGNTLIDLSNNKLDGRIP 719  
 Db 529 ELGHNNLSGPPIWEEFENLKLAVFDLKNALSGSPSSLSGWTSLDALDLSNNRSGSIP 588  
 QY 720 QAMSAITMLEIDLSNNLSGPIPEMGQETFPAPAKLNPGLCG-YLPKCDPBNAGY 778  
 Db 589 VSIQQLSFLSKSVAMNLSGVIPSGGQFQTPPNSSFESN-HLCEHRRFPCSEGTESALI 647  
 QY 779 AHHQSHGRPPASLAGSVAMGLFSPVCFGLIYGR-EMRRKRRKKEALEMYAEHGN 837  
 Db 648 KRERRSRG-----GDIMAGIAGVSFLLTLISLVLRARRSGEVDPEI----- 693  
 QY 838 SGORTNNNTWKLTGYKEALSINLAFEKRLKLTFFADLLQATNGFHNDSLISGSGFGDY 897  
 Db 694 ---ESESNNRKLQ--ELGSKLVVLFQNDKEISVDLLDSTNSPDQANIIIGCGGFGAV 748  
 QY 898 YKAILKQSGAVAIKLIHVSQGGREPMAMETIGIKIRKRVPLLYGCKVGDRLVNE 957  
 Db 749 YKATLPDQKKVAILKLSGDCQIEREFEEVEVLISRAQHPNVLVLEFGCFYKNDRLIIS 808  
 QY 958 VMKYSLEBDLQDPKKGVVYLKSTRKTAIGSARGLAFLHNSCPHIIHRDMKSSNVLL 1017  
 Db 809 YMENGSLDYMLHEHNDSPALLKWKTRIRLQGNAKKELVYHEGCDPHILHRDIXSNILL 868  
 QY 1018 DENLEAVSPFGMARLMSANDTHLSVTLTGTGYVPEYQSPFCSTKGDVSYGVVLL 1077  
 Db 869 DENNSHLADPGLARLMSPEYTHVS-TDLVGTGLYIPPEYGAQSVATYKGDVYSFVGVL 927  
 QY 1078 ELLTGKPTD--SPDPGDNNLVGM-VKQNAKALISVDPEDELMKEDPALTEIELQHLKVA 1134  
 Db 928 ELLTDRKPVDMCKPK-GCRDLISVWVVMKHSRASEVDFPLITSKEN-DKENFRVLEIA 984

Qy 1135 VACLDRAWRPMTYVAMFKEI 1158  
Db 985 CLCLSENKQRPPTQQLVSWLDDV 1008

## RESULT 8

E86308  
hypochemical protein F20D23.7 - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001  
C/Accession: E86308  
R/Theologas, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Hughes, M.K.; Comol, L.; Conway, A.B.; Conway, A.R.; Creeasy, T.H.; Dewar, K.; Chiu, C.W.; Chung, B.; Hutzler, L.  
Nature 408, 856-820, 2000  
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maltli, R.; Marzall, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, K.; M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A/Accession: E86308  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-1133 <STO>  
A/Cross-references: GB:AE005172; NID:g5734762; PIDD:AD50027.1; GSPDB:GN00141  
C/Genetics:  
A/Map position: 1

Query Match 22.1%; Score 1367; DB 2; Length 1133;  
Best Local Similarity 32.9%; Pred. No. 7,8e-55;

Matches 398; Conservative 179; Mismatches 455; Indels 176; Gaps 36;

Qy 8 FLAVTTLFFSPFSLFQASPSQSLYREIHQIISFKDVLPRDKLPPDMS-NKNPCTFD 65  
Db 7 FLAVTTLFFSPFSLFQASPSQSLYREIHQIISFKDVLPRDKLPPDMS-NKNPCTFD 65  
66 GVTGCRDCK-VTSIDSSKPLVGFSAVSSLSLGLSLPLSNHNGSV-SGPKCSAS 123  
Db 60 GIACHTLRLVTVSDANGMUS---GTLSPILCKLHKLKLVSTNFGPPODLSLCS 116  
Qy 124 LVSIDLSRNSLSG-----PYTTLTSL-----GSCGSLKPLVANSNTLD 161  
Db 117 LVLVLDLCTRFHGVPIQLTMTITLKKLYLCENYLFSGIPROIGNLSLQGLVYSNNL 175  
Qy 162 PRGKVSQGL-KXNSLEVLDLSANSISGANVYGVWSDCCGELKHLAISGNKISGV--DV 218  
Db 176 -TGVPSPMAKLRQRLIRAGNGFSG--VIPSEIS-GCESLKYLGAEVLLSEGLPKQL 231  
Qy 219 SRCVNLLEFLDVSSNNFSTGI-PFLGDCSALQHLIDISGNKLSGDSRAISCTELKILNIS 277  
Db 232 EKLQNLVTLIMQNLQSLBEIPSVANISRLVLAHENVYFGSLPRIGIKLTKKRLVLY 291  
Qy 278 SNQGFPIPLP-PLKSIQYISLAENKFTGIP-DPLSGACDTLTGLDLSGNHYGAVPP 334  
Db 222 TNQNLGELPREIGNLIDAAEDIFENQLTGFIPEK--GHILNKLHLLENLILGPIR 349  
Qy 335 FFGSCSLLESIALSSNNFSGELPMDTLLKMGKLVLDLSNFBEGELPESLTVLSASLIT 394  
Db 350 ELGELTLEKLDLSINRANGTIPQ--LQFLPYLDLQFLNQLGKLP-PLIGFYSNFSV 407  
Qy 395 LIDSSNNSFGILPNTLCONPKTLQELVYLQNGFTGKIPTLNSCELSVLHSPVYLSG 454  
Db 408 LDMNSNISGPIPAHFCR--FQTLILSLSGNKLSGNIPRDLTCKSLTGLMGDNLGLT 465  
Qy 455 TTPSISGLSLRLDKLMLNLEBEIPOELMYVKTLETLILDNDLTLGELPSGLSCTNL 514  
Db 466 SLPTLEFLQNLVLALEHQNWLISADLGLKLNLERLANNFTEGIPRIGNLTKI 525  
Qy 515 NWISLNNRLTGLTIPKVIKGLLENLAIKLSNNSPSGNIPELDCSLWLDLNTWLFNG 574  
Db 526 VGFNISNQLTGLHPRKELSGCVTIGRLDLSGNKFSGLYIAQELQGLVYLETILRLSDRLTG 585

Qy 575 TTPAMFKOSKILANFIAGKRVYIKNQDKKRGAGNLEFGGISEQNLRLSTNP 634  
Db 586 EIP-----HSFGDL-----TR-- 596  
Qy 635 CNITSRVYGHTSPTEPDNNSMFLDMSYMLSGYIPKEIGSMPLFI-LNLGNHNDISG 693  
Db 597 -----LMEIQGNLSENIPELGLTSLQSLINSHNLSGT 635  
Qy 694 IPPEVGDRLNLTLSNKLDRIPQAMSALTMLTEIDLSNNLSGPIPEWGFETPP 753  
Db 636 IPDLSNLTQWLETLVNDKLSGEIPASIGNLSLILCNISNNLVGVPTPAFQRDS 695  
Qy 754 AKLNPNGCGVLEPCDP--SNADGVAMHQRSHRRPASA-----GSYAMGLLSFV 805  
Db 696 SNRAGNHGLCNSQRSHCPVLVHSDSKMLWNLINGSOKOKITITCIVISV---FLITFL 752  
Qy 806 CIPGLILVGEREMKRRKKKEALEMYAGHNSGDRNTANTMKTGVKEALSINLAPE 865  
Db 753 GLCWTI-----KREPAFVLE-----DQTKEDV-----MDSY 781  
Qy 866 KPLKLTFLADLQATNGFRNDSLISGFGDYVYKALKDSAVAIIKKLHVSGG--DR 922  
Db 782 FPKGFTYGLVADATRNPSBDVLGRGACGVYKAEVMSGGEVIAVKTL-NSRGGASDN 840  
Qy 923 EPMAEMETIGIKHNLVPLGCKVGERLTVNEMKYSLSEVLDODPKKGVKLTST 982  
Db 841 SFRALSTIGKIRHNIYKLVGFCRQNSNLLTYEWSKSLQDLRGKNCIL-DWNA 899  
Qy 983 RKRLAIGSARGLAFLHNHCSPIIHRDKSSNVLIDENLEARYSDFGMARLMSAMDTLS 1042  
Db 900 RRIALGAAEGLCYHHCRCPOIVARDIKSNLLIDRRFOAHVQDPLALCID-LSTSKS 958  
Qy 1043 VETLGTGVPYPEYVYOSFRCSYGVYVYVLELTGKRPDSDFPDNNLVGVWVQ 1102  
Db 959 MSAAVAGSTGYLAPEYAVMKTEKCDIYSFGVYVLELTITGKPPVQPLEQG-GDLVWVVR 1017  
Qy 1103 HAKLRIS--DVDPPELMKEDPALIEILQHKVVACTLDRAWRPMTYVAMFKEIQ 1160  
Db 1018 SIRNIPITLMEFDARLDNDKRTVHEMSLVKLALFCTSNPSPSRPFRVAVMTTEARG 1077  
Qy 1161 GSGIDSSQ 1168  
Db 1078 SSSLSSS 1085

## RESULT 9

B84742

probable receptor-like protein kinase [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C/Accession: B84742  
R/Li, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Nause, D.; Nierman, W.C.; Cronin, L.A.; Shen, W.; VanAken, S.E.; Umayam, L.; Tallon, L.C.; Nature 402, 761-768, 1999  
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A/Accession: B84742  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-1124 <STO>  
A/Cross-references: GB:AE002093; NID:g2924777; PIDD:AC04906.1; GSPDB:GN00139  
C/Genetics:  
A/Map position: 2

Query Match 22.1%; Score 1364; DB 2; Length 1124;  
Best Local Similarity 33.1%; Pred. No. 1.1e-54;

Matches 395; Conservative 177; Mismatches 483; Indels 138; Gaps 34;

Qy 2 KTFSSFLSTVTLFFSPFSLFQASPSQSLYREIHQI-----SFKVLDKXLLPWS 56  
Db 1 KTFSSFLSTVTLFFSPFSLFQASPSQSLYREIHQI-----SFKVLDKXLLPWS 56



Db 10 KESKMEVGV-----LFLTLVVTSESLSNDGQFILELKNRGFODSL--NRILHNWN 59  
 Qy 57 S-NKNDCTDGTCTRDCKYTTIDLSKPLNNGFSANVSSLSLTGLESFLNSHINSV 115  
 Db 60 GIDETPCNMIGVNC-----SSQ-----GSSSSNSLV----- 86  
 Qy 116 SGFKCASLTSLSRNSLSGPVTTLSLGSCSGLKFLVNSNTL--DPPKVGSGGLKLN 173  
 Db 87 -----VTSLSLSMNSLSGIVS--PSIOGLVNLVYLMALVAGLGDIPREIG--NCS 133  
 Qy 174 SLEVLDSANSISGANVGVWISDGGELKHLAISGNKISGDV--DVSCKVLEFLDVS 231  
 Db 134 KLEVMFLNNQFGGSIIVE---INKLSQLRSFNI CNNKLSGPPREIGDLVNLBELVAYT 190  
 Qy 232 NNSTGTP-FLGDCSALQHLDISGNKLSGDFSAISTCTELKLNLSNQFVPIPL-- 288  
 Db 191 NMLTGLPRLSILGNLNTTTRAGONDPSGNIPTREIGKLNKLLGLAQNISELKEIG 250  
 Qy 289 PLKSLQVLSAENKFTGEIPDPLSGACDTLTGLDLSGNHYGAVPPFSGCSLLESIALS 348  
 Db 251 MLVLTQVILMOKNFSGFTPKDI-GNLTSETLALYGNLSVGPISPEIGNMKSLKLYLY 309  
 Qy 349 SNNFSELPMDTLKMGKLYLDSFNEFSGELPESTLNSALTLTLDSSNFSGPILP 408  
 Db 310 QNOLNGTIPKE-LCKLSKWEIDFSENLISGEIPELSEKIS-ELRLYLQNKLTG-IIP 366  
 Qy 409 NLGNPKVLTQELYLQNNNGFTGKIPPTLSNCELSVLSLFSNYLSGTIPSSLSLSTLRD 468  
 Db 367 NELSKLRN-LAKDLISLNSLTGPIPPGQNLTSMRQQLPFRNSLSGVILPQDLGYSPLW 425  
 Qy 469 LKLMNMLEGIPOLMAYKLTLETLIDFNDLTGEIPSGLSNCTNLWISLNNRLTGEI 528  
 Db 426 VDFSNQSGKIPPICOOSMLILNLGNSNRIFGNIPGVLRCKSLQLQVRVGRGLTGQF 485  
 Qy 529 PKWIGLENLALIKLSNNSFSGNIPDELGDORSLIMDLNTNLFNGTIPAMFROSG-- 585  
 Db 486 PTELCVNLVSLTELQNRFSGPPELGTCKLORLHLAANOSSNLPNISKLSMLVT 545  
 Qy 586 -KIANFLAGKRYVYIKNDGKKECHGAGNLEFQIRSEQLNRLSTRNPNCTISRYVG 644  
 Db 546 FNVSNLSLTGPIPSRIANCKMLQRLDSRN--SFISGLPELGLSHQLEIRLSENPFSG 603  
 Qy 645 HTSPFDNNGSMFLDMSYNNLSGYIPEKISGMPYLF-LNLGNHDSGSI-PDEVGDLRG 703  
 Db 604 NIPFTIGNLTHLTLELQMGNLPSGIPPOLGLSSLQIAMLNSYDFSGEIPREIGNHL 663  
 Qy 704 LNIILDSNKLGRIPQAMASALTMTLEIDLSNNLSGPIPEWQGFETPPKFLNPPGLC 763  
 Db 664 LMYLSLNNHLSGEIPPTTENLSLGLCNFSYNNLTGQLPHQIQLQNMWTLTSFLGNKGLC 723  
 Qy 764 GYPLPCRDPSNADGYAHQSHGRD--ASL-AGSVANG---LFSFYCLFGLLVGRE 816  
 Db 724 GGLHRSQD-----SHSSPHSISLKAASARRRRIIIVSVYIGISILLIIV 772  
 Qy 817 MKRRRKKALEMAEAGHNSGDRANTNNTKLTGVKALSINLAPEKPRKLTPLADL 876  
 Db 773 VHELRNPEVETAYVD-----KAPFQESDIYVPERFTYVDI 812  
 Qy 877 LQNTNGFHNDSLSGSGFVDYKAILKDSAVAIAKL-----IHSVQDGRFEMAEME 929  
 Db 813 LEATKGFHDSYVGRACGVYVYKAWPBGKTIAYKLTJESNRGNNSNNTNDSFPAETL 872  
 Qy 930 TTGIKIRNLVPLLGYC--KVGDELLVNEVMKYSGLEVDLPKKGKGYKLTSTRKTA 987  
 Db 873 TLCKIRIRNLVRLYSFCYHQSNSNLLLYEYMSRGLIGELHGGKH--SMOMPTRFAIA 930  
 Qy 988 IGAAGIATFHNSCPHIIHRDKSSVLLDEMLARVSPFGMARLSAMDTHTSVSTLA 1047  
 Db 931 LGAAGIATYHDOCKPRIIHRDIKSNMILIDENFEAVGFGIAXIID-MPLSKSASAVA 989  
 Qy 1048 GTFGYVPEPYOSFRCTKGDVSYGVVLLLELTGKRPTDSPDGNNLVGVYKQAK-- 1105  
 Db 990 GSYGIATPEAYIMKYTEKCDIYSFGVLLLELTGKAPVQPLEG--GDLATWTRNHIRDH 1048

Qy 1106 LRISDVFDEPLMK-EDPALETELQHLKVAVCLDDBAMRPTMVOVMAFKE 1157  
 Db 1049 SLTSEILDPYLTKEVDVILNMTVTYKIAVLCTKSSPSDRPIMREYVLMLE 1101

RESULT 10  
 705898  
 Hypothetical protein FEH11.170 - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 09-Apr-1999 #sequence\_revision 09-Apr-1999 #ext\_change 22-Oct-1999  
 C/Accession: T05898  
 R/Beyan, M.; Brandt, P.; Dose, S.; Jarke, D.; Scharfe, M.; Schon, O.; Bancroft, I.; Mewer  
 submitted to the Protein Sequence Database, April 1998  
 A/Reference number: 215456  
 A/Accession: T05898  
 A/Molecule type: DNA  
 A/Residues: 1-1003 <BEV>  
 A/Cross-references: EMBL:AL021684; GSPDB:GN00063; ATSP:FEH11.170  
 A/Experimental source: cultivar Columbia; BAC clone FEH11  
 A/Genetics:  
 A/Map position: 5  
 A/Map: ATSP:FEH11.170  
 A/Introns: 867/1  
 C/Superfamily: protein kinase Xa21; leucine-rich alpha-2-glycoprotein repeat homology; p1

Query Match 22.0%; Score 1357.5; DB 2; Length 1003;  
 Best Local Similarity 33.0%; Pred. No. 1.6e-54;  
 Matches 384; Conservative 177; Mismatches 382; Indels 221; Gaps 34;

Qy 14 IFFSFPSLSFQASPSQSLYREIHQSLFQVLT---PDKYL-LPDMSSNKPCTPGVY 68  
 Db 7 LFLFIHISHTFASRPIS---EPRALISLKTSLTGAGDDKNSPLSMKVSSTFTWIGVT 63  
 Qy 69 CRDCK--VTSIDLSKPLNNGFSANVSSLSLTGLESFLNSHINSVGFKCASLTS 126  
 Db 64 CVSRRHVTSIDLSG-----LNLSTLSVDVSHRL-----LQN 97  
 Qy 127 LPLSRSLSGPVTTLTSLGSCSGLKFLVNSNTLD--PPKVGGLKLNLSLEVLDSANS 184  
 Db 98 LSLAENLISGPIP--PEISSLSGLRHLNLSNNVFNFGSPDEISGLV--NLRVLVDVYNNN 153  
 Qy 185 ISGANVGVWISDGGELKHLAISGNKISGDVAVRCVNLBELDVSSNNSFGTGFELDC 244  
 Db 154 LTGDLPLVS---VTNLQTLRHLHAGNYFAGKIP-----PSYGSW 189  
 Qy 245 SALQHLDISGNKLSGDFSAISTCTELKLNLS-SNOFVGPIPL--PLKSLQVLSAEN 301  
 Db 190 PVEYLAVSQNELVWKIPPEIGNLTTLRELTYGYNAFBDGLPPEIGNLSLVRFDGANC 249  
 Qy 302 KFTGEIPDFLSGACDTLTGLDLSGNHYGAVPPFSGCSLLESIALSSNFSGELPMDL 361  
 Db 250 GLTGEIPPEI-GELQTLDTLFLQVNVFSGPLTWELGTLSLSMDSNMFTEGELPA-SF 307  
 Qy 362 LKMGKLYLDSFNEFSGELPESTLNSALTLTLDSSNFSGPILPNCNPKVLTQEL 421  
 Db 308 AELKNTLILNLPNNKJLHGEIPPEIGDL-DELEVLTQIMENNFGSIPQKLGNGKLNLYD- 365  
 Qy 422 YLQNNNGFTGKIPPTLSNCELSVLSLFSNYLSGTIPSSLSLSTLRDLKLMNMLEGIP 481  
 Db 366 -LSSNKLTLPLPMMCSGNLTLETLITLGNLPSGISDQKCSLRIRIMGENFLNGSIP 424  
 Qy 482 QELMYVKTLETLIDFNDLTGEIPSGLSNCTNLWISLNNRLTGEIPKMGIGLEMLALI 541  
 Db 425 KGLFGLPKLTQVLELDVYLSGELPVAAGVSVNIGQISLNNQISGGLPPAIGFTGQKL 484  
 Qy 542 KLSNNSFSGNIPDELGDORSLIMDLNTNLFNGTIPAMFKQSGKIAANFIAGKRYVYIK 601  
 Db 485 LLDNKKFGPIPSFVG----- 500  
 Qy 602 NDGKKECHGAGNLEFQIRSEQLNRLSTRNPNCTISRYVGHTSPFDNNGSMFLDM 661  
 Db 501 -----KLQQLSK-----IDF 510



QY 662 SYNM:SGYIPKEIGSMPLFLILN:GNDISGIPDEVDLGLNILDLSNKLDRIPQA 721  
 Db 511 SHNLSGIGIABEISRCXKLLTFVDLSRNLSEGISPEIETAMKILYNLSRNLVGSIPGS 570  
 QY 722 MSALTMLEIDLSNNNLSGPIPEMGOFETPPAPKLNPGICGYPLPRCDSDNADYAH 781  
 Db 571 ISSMOSLSLDPSTYNNLSGLVPGTQSFYNTSFLGPNDLCCGYPLGCKDGVAKG-GHQ 629  
 QY 782 QRSHGRRPASLAGSVAMGLFSPVFCIGLLVGRBMRKRRKKEALEMYAFGHGNSGDR 841  
 Db 630 SHSKGPIELASMKLLVLIGLL---VCSIAFAVVA-IIRKSLKXSE----- 671  
 QY 842 TANTNMKLTGVKELSLNLAFAEKPLKLTFADLLQATNGFNHDSLISGGFGDYKAI 901  
 Db 672 ---SRAMKLT-----AFQR-LDFTCDVL---SLKEDNIIIGKAGAIYKGV 712  
 QY 902 LKDGSAVAIKKLIHVS---GGQDRFEMAMETIGIKIRNLVPLIGCKVDERILNVEM 959  
 Db 713 MPNDOLVAVKRLAMSGSSHDHGFNAEIQTLGRIRHRIHVLRFCSNHEHTNLVYEM 772  
 QY 960 KYGSELDVLDQPKKGYKLLKSTRKIAIGSARGLAFLHNHCSPHIHRDKMSNVLLDE 1019  
 Db 773 PNGSLGEVLHG-KKGG-HLHWDTRYKIALERAKGLCYLHDCSPLIYHRDVKNNILDS 830  
 QY 1020 NLEARVSDPGMARLMSAMDHLSTLACTPGYVPEPYTOSFRCTSTGYDYSGVILLET 1079  
 Db 831 NFEHVAHDFGLAKFLQDSGTSECMASALASGYIAPAYATLLKVDKESDYVSGVILLET 890  
 QY 1080 LTRKRPDSDPDGDN-NLVGMVYKQHA---KLRTSDVPDFELMKEDPALTEIHLQHL-KYA 1134  
 Db 891 VTGKRKPYG--EFGDGVLDVGMVNRMTDSNKDSVLKVLDPNL---SSIPHEVTHAVYA 944  
 QY 1135 VACLDPRAMRPTMVQVAMFKKI 1158  
 Db 945 MLCVBEQAVRPTMRREVYQILTEI 968

RESULT 11  
 967746  
 hypothetical protein T9N14.20 [imported] - Arabidopsis thaliana  
 C:\Species: Arabidopsis thaliana (mouse-ear cress)  
 C:\Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 A:\Accession: G967746  
 R:\Theologos, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Com, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 Hansen, N.F.; Hughes, B.; Hultzer, L.  
 Nucleotide 408, 816-820, 2000  
 A:\Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:\Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,  
 Ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.  
 A:\Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:\Reference number: A86141; MUID:21016719; PMID:11130712  
 A:\Accession: G967746  
 A:\Status: preliminary  
 A:\Molecule type: DNA  
 A:\Residues: 1-1095 <STO>  
 A:\Cross-references: GB:AE005173; MID:g10645385; PIDN:AA621504.1; GSPDB:GN00141  
 C:\Genetics:  
 A:\Gene: T9N14.20  
 A:\Map position: 1

Query Match 21.8%; Score 1349.5; DB 2; Length 1095;  
 Best Local Similarity 32.4%; Pred. No. 4.7e-54;  
 Matches 395; Conservative 192; Mismatches 398; Indels 233; Gaps 42;

QY 3 TFSSEFLSYT-----TLFSPFSLSQASPQSLVREIHQLSFCDVLPDKLL 52  
 Db 35 SLSVFFLVYSEAVCNLQDRDSLWMS-GNVSSPVS-----LH----- 71  
 QY 53 PWSNKNKPCITGVTC---RDDKVTSIDLSKPLNVGFSAVSSSLSLTGLSESLFNS 109

Db 72 --WNSSIDCCSMGISCDSPENRVTSIILSRGSL---GNLPSVLDLQRLRIDLSLN 126  
 QY 110 HINGSV-SGEKCSA-SLISLDSRNSLSGPVTTTSLGSCGLKELFANVSNTLDPGKVS 167  
 Db 127 RLSGLPFGFLSALDQLVLIDLISYNSFKGELPLQSPGN-----GNNI-PP----- 172  
 QY 168 GGLKNSLEVLDSNLSISGANVGVWLSDCGGLKHLAISNKSIGDYVSRVNLLEPL 227  
 Db 173 -----IQTVLDSN-----LLEGEIILSSVPLQGANLTSF 203  
 QY 228 DVSSNNFSTGIP-FLDGSA---LOHLDSGNKLSDGFSRAISTCTELKLNISNGV 283  
 Db 204 NVSNNSFTGISFWM---CTASPQLTKLDPSTYDFSGDLSQELSRCSRLSVRAGNNLSG 261  
 QY 284 PLPP--LPLKSLQYLSLANKETGSIPLPLSGACDTLGLDLSGNHFYGAAPPFGCSL 341  
 Db 262 ELPKEIYMLPELEOLFPLPNRLSGKIDNGIT-RLTKLTLLEYSHHIGEIPKDIGKLSK 320  
 QY 342 LESLALSSNNFSGELPMDTLKMRGLKVLDSFNEFSGELPESLTNLSASLITLDSNN 401  
 Db 321 LSSLQHLVNNLMGSLPV-SLANCTKLVKLNLRVNLQGLTSLAIDFSRQSLSIDLG--- 376  
 QY 402 FSGPILPMLCONPKNTLQELYLONNGFTKIPPTLSNCSFELVSLHSFNYLSGTIPS--- 458  
 Db 377 -----NNSFTGEFPSTVYSCKMTFARFAGNKLTLQSLSPVL 413  
 QY 459 SLGSL-----KLRLKMLNMLEGEPDOLMYVKTLETLLDPDLGEIPSG----- 507  
 Db 414 ELESLSFTFSDNKNKNTNLGALSLIOG-----CKLSTLIWAKNFYDETVSNDFL 465  
 QY 508 -LSNCTNINWISLSNNRLTGEIPKMWIGLEMLALIKLSNNSFSGNIDELDCSLTMD 566  
 Db 466 RSDGPPSLQIRGICACRLTGEIPAMLIKQVREWMDLSMRPFVTIGWGLTLPDLFLD 525  
 QY 567 LNTNLFNGTIPAMFKOSGKIAPNFIAKRYVYIKNDGKKECHAGANLLEF-----Q 619  
 Db 526 LSDNFLTGELEPKELFOLRA-----LMSQKAY-----DATER-----NYLELPVAVPNN 569  
 QY 620 GIRSEOLNRSLSTRNPNCTITSRYVGHSTSPFPDNNGSMFLDMSYNNLSGYIPKEIGSMY 679  
 Db 570 VTTNQQNVQLSLPPTYIKR-----NNLTGTIPVEVGQLV 606  
 QY 680 LFLINLGHNDISGIPDEVDLRLGILNLDLSNKLDRIPQASALTMLEIDLSNNLS 739  
 Db 607 LHIELLGNNSGSSIPBELNLTNLELIDLSNNLSRLIWSLGLHFLSYEVANNTLS 666  
 QY 740 GPPEMGQFETPPAKFLNPNGLC-GYPLPRCDPSNADGYAHHQSHGRRPASLAGSVAM 798  
 Db 667 GPITGQFDTFPYANFEGNPLLCGVLTLSQDPTQ---HSTTKMGKGYKNRTLVLGLVL 723  
 QY 799 GLTSPFCIGLLIVGEMKRRR-----KKEALEMYAEGH-----GNSGRTYANNTN 847  
 Db 724 GLFEGVLI--LVLLALVLSTKRRVNDGSENALEINSGSYSEVPAGSDKD----- 774  
 QY 848 WKLTVKEALSLNLAFAEKPLKLTFADLLQATNGFNHDSLISGGFGDYKAIKDGSA 907  
 Db 775 -----ISLVLPFGNSRE--VKDLITFELLKADNSQANIIICGFGVLYKATLNGTK 827  
 QY 908 VALKLIHVSQGQDRFEMAMETIGIKIRNLVPLIGCKVDERILNVEMKYGSLIEDV 967  
 Db 828 LAYKQLTGDYGMKEKEKAEVETLSRAKHENLVAALQGYCVHDSARILIYSFMENSGLDYV 887  
 QY 968 LQDPKKGVLLKLTSTRKIAIGSARGLAFLHNHCSPHIHRDKMSNVLLDENLEARYSD 1027  
 Db 888 LHNHPEPQALDMPKLNLMKAGSSGLATMHCPEPHIYHRDKSSNILLDNFAYVAD 947  
 QY 1028 FGMARLMSAMTHLSVSTLAGTPGYVPEPYTOSFRCTSTGYDYSGVVLLETLTKRAPD 1087  
 Db 948 FGLSRLLPFRYHTV-TELVTGLGYIPREYQAMVATLRGDVYSFGVWMLLETLTKRPM 1006  
 QY 1088 --SPDGDNNLVGMVYKQHAKR---ISDVPELMEKEDPALTEIHLQHLKVAVACLDRA 1142

Db 1007 VPRPKM-SRELVAHV--HTWKRDGKEEVED-TILRES-GNEBAMLRLVLDIACMVNQN 1061  
 QY 1143 WRPPYVQVMAFKEIOA 1160  
 Db 1062 MKRPNIQVVDWMLKNIEA 1079

## RESULT 12

hypothetical protein F23E13.70 - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 20-Sep-1999

C/Accession: T04587

R/Author: M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Jesse, T.

submitted to the Protein Sequence Database, March 1998

A/Reference number: 215378

A/Accession: T04587

A/Molecule type: DNA

A/Residues: 1-1134 <REV>

Experimental source: cultivar Columbia; BAC clone F23E13

C/Genetics:

A/Map position: 4

A/Intons: 960/3

A/Note: F23E13.70

C/Superfamily: protein kinase Xa21; leucine-rich alpha-2-glycoprotein repeat homology; F

Query Match 21.7%; Score 1340.5; DB 2; Length 1134;

Best Local Similarity 31.7%; Pred. No. 1,36-53; Indels 217; Gaps 38;

Matches 396; Conservative 190; Mismatches 447;

9 LSVTLFFSPFSPSPQSPQSLVREIHQISFKDVLDPK-NLLPDW--SSNKNPCTFD 65

3 ISLFFPLVIVAPLVSYADESGA--EIDALTAFLNHDPLGALTSWDPSTPAFCDMR 59

66 GYTCGDDKYSIDLSKPLNVSFAVSSSLSTGLTESLFLNSHINGSVSGFCSASLT 125

60 GVCCTNHRVTELRPR-----LQLSG-----RISRIISGLR---MLR 93

126 SLDDLRSNLSGCVTLTISGSCGKFLNVSNTLDPPGKVSGLK-LNSLEVLDSLNS 184

94 KLSLNSNSNG--TIPSLAYCTRLLSVFLQVNSLS--GKLPPAMRNLTSLEVEN---- 144

185 ISGANVAVGLSDGCGELKHLAISGNKISGDVDSRCVNLFLDVSSNNFSGIPLDGC 244

145 -----VAGNRISGEIPLVGLPSSLQFLDLSMTFSGQIP----- 177

245 SALQHLIDISGNKLSGDFSAISTCTELKILNINSNOFVPIPL--PKSLQYLSLAENK 302

178 -----SG-----LANTQLQLNLSYNQLTGEITASLGNLQSLQYMLDPLNL 219

303 FTEGIDPFLSGACDITLGLDLGNNHFYAVPPFPFSGSLLLESIALSSNNFSGELPMD--- 359

220 LQCTLPFSAISN-CSSLVHLSASENEIGVYPAAYGALPKLEVLSSNNFSGVFPFSLFC 278

360 ----TLTKM-----RGLKVLDSFNFGSGELPESLNLNLSASLTLTDL 397

279 NLSLITVIOGFNAFSDIVRETTANCRGTQVLDQENISGFFPMLNLT--LSLKNLDV 337

398 SSNNFSGPILPNLCONPKVTLQELVYONNGFTKIPPTLSNCELSVLSHLSFVYLSGTP 457

338 SGNLFSGEIPLDI-GNLR-LLEIKLANSLSIETIPEIKQCGSLDVLDFEGSLKQGIIP 395

458 SLSGLSKLDLKLMLNMLEGEIIPQELMVKYLTETLLDPNDLTGEIPSGLSNCTNLMNI 517

396 EPLGYMKALKVLSIGRNSFGVYPSVSVNLQQLERLNLGNNNGSFPVELMALVLSLSEL 455

518 SLSNNRLTGTEIPKWIQLEMLALIKLSNNSFSGNIPDELGDCKSLMLDNLTLFNGTIP 577

456 DLGNNRSGAVPVSISNLSNLSFNLISGNGSGEIPASVGNLFLKTLALDLSKQMGGEVP 515

578 AAFKQSGKIAAMFIACKRYVYIKNDGKKCHGAGNLTLEFQGISRSHQNLRLSTNNPCNI 637

Db 516 VEL---SGLPNVQVIALQ-----GN--NFGSVNDEGSSLSVLSRYNLI 553

QY 638 TSVRYGHTSYTF-----DNN-----GSMVPLDMSYMLSGYIPKE 673

Db 554 SSNSFGEIPQTFPGLRLVLSISDNHISGSLPPEIGNSALEVLELRNRLMGHLPAD 613

QY 674 IGSMVPLFIINLGHNDISGSLPDEVGDLRGILNLDLSNKLDRGIPQAMALTMLTEIDL 733

Db 614 LSLRLKLVLDLQNNISGELPPEISQSSLSNLSLDHNLGSLVPSFGLSNLTQMDL 673

QY 734 SNNMLSGP-----EMGQETPPA-----KFLNPPGLCGVPL- 767

Db 674 SYNNLTGEPASLALISNLYFVNSNNLKGEIPALISGRINNTSPSGNTETLCGPPLN 733

QY 768 PRCPDSNADGYAHNORSGRRPASLGSVMWG-LTFSFVCIFFGLVIGREMRKRRKTEA 826

Db 734 RCESTTAEG-----KKQRKMLIYMAALGALLSLFCFYYTTLTKRKTKLKQOSTT 788

QY 827 ELEMYAEGHNSGDRNTANNMKLT--GVKEALSLNLAEPKPLKLTLPADLQATNGFH 884

Db 789 GKRSRPTAGSRVRSSTSRSTENGEPKLVFN-----NKITLLETIATROFD 840

QY 885 NDLISGGGFEDVYKAILIKGSAVAIKLIVHSGQGRFEMAEETIGKIGRNLVPLIG 944

Db 841 EENVLSRTTYGLFPRANYNDGVLSIRLPNGSLNLENLFKKEABVGLKVRHNIYVLRG 900

QY 945 -CKYKGDRLVNEVMKGSLEDVLDQPK-KGVVTKLSTRKIALISGARGAFLHNCS 1002

Db 901 YVAGPDLRLVYDMPNGNLSLTLQASHODGHVLMNPMHLLALGARGLGLHQS-- 958

QY 1003 PHIIHDMKSNVLLDENLEARVSDFGMARLMSANDHLSV-STLAGTPGYVPEYYSF 1061

Db 959 -NMVHDIKQVVLDPADPEAHISDFGLRLIRSPRSAAVTANTIGLVYSPPEATLSG 1017

QY 1062 KCTKGDVSYGVVLELLTGKRPDSDPFGDNNLVGVV-KQHAKTARISDVEPFLMKED 1120

Db 1018 EITRESDIYSFGVILEITLTKRPVWFTQ--DEDIVKWKQKLOQGVYTELLEPFLBED 1075

QY 1121 PALE--TEYLQHKVAVACLDRAPRPTMVQVMAFKEIQAGSGIDSQS 1168

Db 1076 PESSEWEFLLIGIKVGLCTATDPDLPDPTMSDVVFMLECCRGVPPVPSGA 1125

RESULT 13

probable Protein kinase [imported] - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C/Accession: B86465

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

R/Author: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

```

Db 6 SNFF-----FLFPCSVSMAPQTLTSLSSGQALSLK--RSPSLFSSMDQDQTPCS 57
Qy 64 FDVTC-RDDKVTISIDSKPLNVFSAVSSSLTGLTSLFISNHNIGSV-SGFKKS 121
Db 58 WYGITCSADNRVIVSIPDFLNL--SSI-PDLSLSSLOFLMSSNNISGPIPPSFGKL 114
Qy 122 ASLTLSSRLSGPVTTLTSLGSCGLKFLNVSSNTL--DPPKAVSGGLKLNLEVL 179
Db 115 THLLDLSSNLSGPIIP--SELGRSLTLOPLILMANKLSGISIPQIS--NLFALOVLC 169
Qy 180 LSANSISGA--NVGWVYLSDDCGELKHLAISGNKISGDVDVSCVNLLEFLDVSSNNFSG 237
Db 170 LQDLMLNGSIPSSFGSLVS-----LQDFLNGN-----TNLGGP 203
Qy 238 IFF-LGDCSALOHLIDISGNKSGDPSRAISTCTELKLNISNQVGPPIP-LPIKS-LQ 294
Db 204 IPALGELKMLTTLGFAASGLSGSIPSTFGNVLNQLMALYDEISGTTIPQLGCLSELR 263
Qy 295 YLSLAERKFTGEIPDPLSGACDITLGLDLSGNHFGYAVPPFGSGCLLSLSSNNFSG 354
Db 264 NLYIANKKLTGSLPKEL-GKLQKITSLILMGNLSGVIPPEISNCSLIVFVDSANDLTG 322
Qy 355 ELPPDITLKNKGLKVLDSLSEFSGELPESLTYLSALITLTLSSNNFSGPILEMLCONP 414
Db 323 DIPED-LGKLWMLBQQLSDMMFTQIPELNSCS-SLALQDLKXKLSGSI-PSQIGNL 379
Qy 415 KNTLQELYLONNGFTGKIPTLSNCSLVSLSHFNYSGLTIPSSIGSLSKRLDKMLN 474
Db 380 K-SLQSFEMENISIGTIPSSFNCTDLVALDLSEKLTGRIPLELFSLKSLKLLGN 438
Qy 475 MLGEIPQELMYKTLLETILIDFNLTGELPSGLSNCNLMNISNNRNLGELIPKMTGR 534
Db 439 SLSSGLPKSVAKQSLVRLRVEHQSLGQIPKEIGLOLVLDLYNHFSGPYEISN 498
Qy 535 LENTLALIKSNSSFGNIPDELGDCCSLIWDLNTNLFGTIPAMFQSGKIANFIAG 594
Db 499 IYVLELDVHNNTITGDIQAQGNLVNLEQDLSSNSFTGNIPLSF-----544
Qy 595 KRVYVYKDNQMKKECHGAGNLEFQGISREQLNRLSTRNPNCTISRYVGGHTSPFDNG 654
Db 545 -----GNL-----SYLANKLIANN-----NLITGQIPKSIKMLQ 572
Qy 655 SMMFLDMSYMLSGYIPKEIGSMPLYFI-LNLGNDISGSI-PDEVGDRLGNTIIDLSSNK 713
Db 573 KTLTLDLSTNSLSEIPQELGQVTLTINLIDSYTFTGNIPETPSDLTQLOSLDSSNS 632
Qy 714 LGRIPQAMSLTMTLEIDLSSNNLSGPIPEMGOFETPPPAKFLNNGLCGYPLRCDPS 773
Db 633 LHGDI-KVIGSLTSLASINISCNNSFGPIPTPFPKTIISTTSYLONTLC-----H 682
Qy 774 NADGY---AHORSHGR-----PASLASVAMGLLFSVCI-FGLILVGREKRRRK 823
Db 683 SLDDGTCSSHTGQNGGVKSPKIVALTAVILASTITAILAANLL-----LENNHLY 733
Qy 824 KEAELMTAEGHNGSGDRTANNNTKLTGVKEALSINLAPEKRLKLTADLQANNGP 883
Db 734 KTSQ-----NSSSPETAEDPSYPTWFI-----PQDLGTT-VANNYVTSL 772
Qy 884 HNDLSIGSGFDVYKALKKDSAVAIKLIHVSQGDG-----EFMAEMETIGIKIR 937
Db 773 TDENVYGGCGSIYKAIIPNGDIAVAVKMKYKDNNEEGESTIDSPABEIQLIGNIRRH 832
Qy 938 NLVPLLYGCKVDERLLAVNEVMKYSLEVDYDPKGGVTLKULSTRKAIISARGIATL 997
Db 833 NIVKLLGYCSKSVKLLIYNFPNGNIQOLLO-----GNRLDMETRYKAIIGAAGIATL 888
Qy 998 HHNGSPHIIHDMKSNVLDLNLNFAVSDFGMARL-MSADTHLSVSTLAGPGVYPE 1056
Db 889 HHDVPAIHHADVKNINLSDSKYEALILADFGAKLMMNSPNYHNSRYA-----E 940
Qy 1057 YVGSFRCTKGDVSYGVVLELTLTGKRPDSDPFGDN-NLVGVYQAHAKLRIISVDPE 1115

```

---

```

Db 941 YGYTMNITEKSDVSYGVVLELTLGSGRAVE-PQIGDGLHIVEMKKK-----MGTPEDA 994
Qy 1116 LMKEDPLF-----TEILOHLKVAACLDRAWRPRPMVQAMFKRIQAS---GID 1165
Db 995 LSVLDVLCQGLPDQIVQEMQLGTIGIAMFCVNSPYERPFMEKVITLNEVKSPEEMKKT 1054
Qy 1166 SQSTIR 1171
Db 1055 SQPLIK 1060

RESULT 14
T10659
Probable serine/threonine-specific protein kinase (EC 2.7.1.1) T5F17.100 - Arabidopsis t
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999
C:Accession: T10659
R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancroft
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z16533
A:Accession: T10659
A:Molecule type: DNA
A:Residues: 1-1013 <BEV>
A:Cross-references: EMBL:AL049917; GSPDB:GN00062; ATSP:T5F17.100
A:Experimental source: cultivar Columbia; BAC clone T5F17
C:Genetics:
A:Gene: ATSP:T5F17.100
A:Map position: 4
A:Insertion: 876/1
C:Superfamily: protein kinase Xa21, leucine-rich alpha-2-glycoprotein repeat homology; p
C:Keywords: phosphotransferase; protein kinase

Query Match
Best Local Similarity 31.1%; Score 1313; DB 2; Length 1013;
Matches 377; Conservative 190; Mismatches 397; Indels 248; Gaps 39;

Qy 9 LSVTLTF-FFSEFSLSPQASPSQSLYREIHQILSPKOVLPDK-NLPPDW--SSNKPCTP 64
Db 3 MKIIVLFYCYIGSTSSVLASINDVNBLSVLSKSTVDPNLFKDKMLSDTSDHCNW 62
Qy 65 DQVTCRD-KYTSIDLSKPLNVGFSVSSSLSLTGLSLFLSNHNGSVSGFKCAS 123
Db 63 TGVRCNSGNVEKDLAAG-----NMLTG-----KIDPSIQGL--SS 96
Qy 124 LTSLDLSNLSGPTVTLTSLGSCGLKFLNVSSNLTLPFGKXVSGGLKANSLEVLDSAN 183
Db 97 LVSEFI-----SCNGFESLIPKS-----IPPLKSIDISON 126
Qy 184 SIGANVGVWLSDCGELKHLAISGNKISGDV--DVSRCVNLEFLDVSSNNFSGTIP-F 240
Db 127 SFSGSL--FLFSNESLGLVHLNAGNNSLGNLTEDLGNVSLVLEVDLDAGNPFQSLPS 183
Qy 241 LGDCSALOHLIDISGNKSGDPSRAISTCTELKLNISNQVGPPIP-LPKSIQYSL 298
Db 184 FKNLOKRFGLSSNNLTGELPSVLQQLPSLETAIIGVYEFKGPPIPERGININSIKYLDL 243
Qy 299 AENKFTGEIPDPLSGACDITLGLDLSGNHFGYAVPPFGSCSLLSLSSNNFSGELPM 358
Db 244 AIGLSEIPSEL-----GKLSLETLLEYENNFTGTTIPR 278
Qy 359 DTLAKMGKLVLSFNEFSGELPESLTNLSASLTLTLSSNNFSGPIPLNLCQNPXYL 418
Db 279 E-IGSITTLVLYLFDSDALTGELIPMEITKL-----KQ-L 310
Qy 419 QELVYONNGFTGKIPTLSNCSLVSLSHFNYSGLTIPSSIGSLSKRLDKLMTMLMEG 478
Db 311 QLNLMANKLSGSIIPALISLAQVLELMNNTLSGEPSDGLKXKSPLOWLDVSSNSPSG 370
Qy 479 EIPQELMYVTLLETILIDFNLTGELIPSGLSNCNLMNISNNRNLGELIPKMTGRNL 538
Db 371 EIPSTLCNKNKNTKLTILFNNTFGQIPATLSTCGSLVRYBMQNNLNSIPIPGKLEKL 430
Qy 539 ALLKSNNSPSGNIPELDDCSLITLWDLNTNLFGNTIIPAMFK-----QSGKIANFIAG 594

```

```

Db 431 QREIAGNLSGIFEDISDSVLSIPDSRQIRSSPSTLSTLHINQAPLVADNFISG 490
Qy 595 KRYVYIKNDGMKKECHGAGNLLFQGISQJNLRSTNPNCTITSRYVGGHISPFEDNNG 654
Db 491 E-----VPDFQOCF 500
Qy 655 SMMFLDMSTYMLSGYIKKIGSMPLFIINLGHNDISGIPDEVDLRGLNILDLSNKL 714
Db 501 SLSTNLDLSNVTLTGTLPSSTASCEKVLSTLNRRNITGTLPIQITMSLAVLDLSNNSL 560
Qy 715 DGRIPQMSALTLTEITDLSNNNLSPIDEMGGFETFPAPKFLNPGLCGYPLPRCDPSN 774
Db 561 TGVLPESIGSPALTELLNNSYNTLTGPVPINGTLTNPDDLKNGSGLCGGVLPCCSKFO 620
Qy 775 ADGVAAHQSHRRPAPS-----LAGSVAMGLFSPFCIGLILVGBEMKRRRKEAELE 829
Db 621 R-ATSSSHSLHGKRIYAGMLIGIASVLTALGI-----LTVITVTLTKK----- 661
Qy 830 MYAEGHNSGDRANTNTNMTLTVGKEALSINLAPEKPLKTLFADLLQATNGFHNDSLI 889
Db 662 WYSNGF--CGDELTASKEMW-----RLMAFHR--LGFIASDILAC--IKESNMI 705
Qy 890 GSGGFGDVYKALIKDGSAY-AIKKL-----IHVSQGDREPAEMETTKIKGRNLVPL 942
Db 706 GMGATGIVYNAEMGRSSTVLAVKLMRSADIEDGTGG--FVGEVNLTKGKHNRNIVRL 763
Qy 943 LGYKRVGDEHLVNEVWKYSGLEVDLDPKKG-VTKLSTRRIATGARGALAFHNG 1001
Db 764 LGFLYNDKNMIVYEFMLNGLGAIHGKNAAGLLVDWISRYNIALGVANGLAYLHDC 823
Qy 1002 SPHTIIRDMKSNVLDLNTENLARYSDFGMARLMSAMDTHLSVSTLACTPGYVPEYQSF 1061
Db 824 HPIYTHDIKNNIILDPANLADARLADFGARMMARKKE--TVSWAGSYGVIAPYCYTL 881
Qy 1062 RCRSTKGVYSGVVLLELTGKRPTDSDPDGN-NLVGMVQKAKLIS--DVEDPELT-- 1116
Db 882 KYDEKIDIVSYGVVLELTGRRPLE-PEFGESVDIYEVWVRKTRDNI SLFEALDPVNG 940
Qy 1117 ---MKEDPALTEILLOHKVAVACLDREMRPFWVQVAMFKEI--QAGSGIDSOSTIR 1171
Db 941 CRYVGE-----EMLVQIALCTTKLPKDRPSRDVISMGLGKPKRKSNSMEENTSR 994
Qy 1172 SIED---GGFST 1180
Db 995 SLAEKHSSVFST 1006
Qy 15
Db 15
S: 36
receptor-like protein kinase 5 (EC 2.7.1.1) precursor - Arabidopsis thaliana
N: Alternate names: protein P2009.180
C: Species: Arabidopsis thaliana (mouse-ear cress)
C: Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 11-Jun-1999
C: Accession: S27756; T04620
R: Walker, J. C.
submitted to the EMBL Data Library, February 1992
A: Description: Receptor-like protein kinase genes of Arabidopsis thaliana.
A: Reference number: S27754
A: Accession: S27756
A: Molecule type: mRNA
A: Residues: 1-999 <MAL>
A: Cross-references: EMBL:M84660; NID:g166849; PID:AAA2859.1; PID:g166850
A: Experimental source: strain Columbia
R: Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hohnle, J.; Mewes, H.W.; Meyer, K.F.X
submitted to the Protein Sequence Database, October 1998
A: Reference number: Z15380
A: Accession: T04620
A: Molecule type: DNA
A: Residues: 1-999 <BEV>
A: Cross-references: EMBL:AL021749
A: Experimental source: cultivar Columbia; BAC clone F2009
C: Genetical:

```

```

A: Gene: RLKS
A: Map position: 4
A: Introns: 868/1
A: Note: P2009.180
C: Superfamily: protein kinase Xa21; leucine-rich alpha-2-glycoprotein repeat homology; p
C: Keywords: ATP; autophosphorylation; glycoprotein; magnesium; phosphotransferase; recep
F: 1-20/Domain: signal sequence #status predicted <SIG>
F: 21-99/Domain: receptor-like protein kinase #status predicted <MAT>
F: 66-89/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR1>
F: 90-114/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR2>
F: 115-139/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR3>
F: 140-163/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR4>
F: 164-187/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR5>
F: 188-212/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR6>
F: 213-236/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR7>
F: 237-260/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR8>
F: 261-284/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR9>
F: 285-307/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR10>
F: 308-331/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR11>
F: 332-355/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR12>
F: 356-379/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR13>
F: 380-403/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR14>
F: 404-427/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR15>
F: 428-451/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR16>
F: 452-475/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR17>
F: 476-499/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR18>
F: 500-523/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR19>
F: 524-547/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR20>
F: 548-570/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR21>
F: 571-593/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR22>
F: 681-971/Domain: protein kinase ATP-binding motif
F: 689-697/Region: protein kinase ATP-binding motif
F: 98-102, 150, 269, 282, 452, 576/Binding site: carbohydrate (Asn) (covalent) #status
F: 711, 737, 819, 821/Active site: Lys, Glu, Asp, Lys #status predicted
F: 824, 828/Binding site: magnesium (Asn, Asp) #status predicted

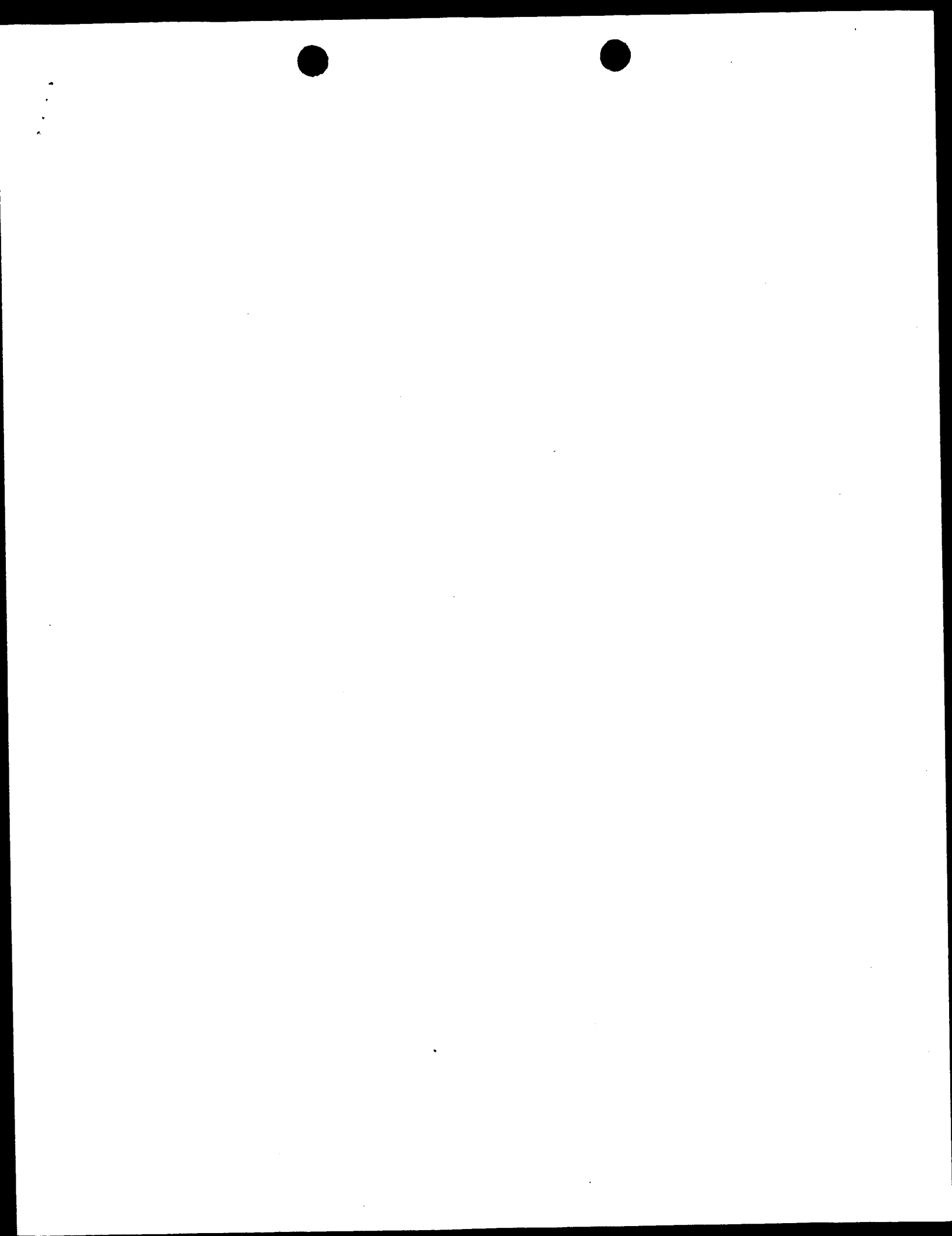
Query Match 21.2%; Score 1311; DB 1; Length 999;
Best Local Similarity 32.4%; Pred. No. 2,46-52;
Matches 386; Conservative 161; Mismatches 392; Indels 254; Gaps 30;

Qy 22 LSPQASPSQSLYREIHOISFEDVLPD-KNLLPMSNKN--PCTFDGVC-RDDKVTSI 77
Db 11 LSTYLPSTLNQDATTILROAKGLSDPAQSLSSMSDNDVTPCKWLGVSODATSNVSV 70
Qy 78 DLSKPLNVFSAVSSLSLTGLSTFLSNHNGSVG--FKCASLTSLDSRNSLS 135
Db 71 DLSFPLVGPFSI---LCHLPSLHSLSYNNSINGSLADDFTCNMLISLSENLV 127
Qy 136 GPVTTITSLGSCGSLKFLNVSNTIDFPKVSGLKANSLEVLDSANSISGANVGVWL 195
Db 128 GSIP-----KSLPFLP----- 139
Qy 196 SDCCGSLKHLAISGNKISGDVDVSRVNLFLPVSNNFSTGIP-FLGDCSALQHLDISG 254
Db 140 -----NKLFLIEIGNNLSDTIPSPFGFRKLSEINLAG 172
Qy 255 NKLSPFRAISTCTELKLNISNOF-VGPIPLP--PLKSIQYSLANKRTGRIPL 311
Db 173 NPLSGTIPSLGAVTTTELKELAYNLFSPQIPSOQGNLTQVYMLACNIVGPIPSL 232
Qy 312 SGACDTITGLDLSGNFYGAVPPFGSCSLSESLASNNFSGELPMPTLLKMRGLKVD 371
Db 233 S-RUTSLVNDLDFNOLTSIPSWITQLKTBOELFNNFSFSELP-BEMGNMTLLKRD 290
Qy 372 LSPFESGELPESITVMSASILTLDSSNNFSPILPNCQNPKNLTQELVYQNGCFQK 431
Db 291 ASNNKLTGKIPDNINL--NLBSLNFENMLBEGPLPESITRS--KTLSEIKLFNNRLTV 346
Qy 432 IPTLNSCSEIVLSHSFNYLSTGTPSSLSLSKADIKLMLMGEIPQELMYKYLE 491
Db 347 LPSQLGANSPLQYVDLSYNFSGEIPANVCGEKLLEYLLIIDSFGESLNNIGKCKSLT 406
Qy 492 TLIDPNDNGEIPSGISNCTNINWISLNNRLTGEIPKWIIGLENIATILKSNNSFSGN 551

```

Db 407 RYALSNKLSGOIPHFWMGLPRLSLLESDNSFTGISPKTIIIGAKNLSNLRISKNRFGS 466  
QY 552 IPDELDCBSLIIMLDLNTNLFNGTIPAMFKQSGKIANFLAGKRYVYIKRDKMKKECHG 611  
Db 467 IPNEIGSLNGIIEISGAENDFSGEIPESL----- 495  
QY 612 AGNLEFGQIBSEQJLNRJSTNPNCTSRVYGHTSPPTDNGSMFMDMSYNNLSGYIP 671  
Db 496 -----VKLKQLSR-----LDLSKNQLSGEIP 516  
QY 672 KEIGSMPLYFIINTLGHNDISGIPDEVDLARGNITLDSNKLDRIPOMASALTMLTEI 731  
Db 517 RELRGKMLNLELANNHLSGEIPKXVGILPVLNYLDLSNQFSGEIPLLEONL-KLNVL 575  
QY 732 DISNNLSGPIPEMGQFEPFPAKFLNNGLCGYPLPCDCPSNADYVHHQSHGRRPAS 791  
Db 576 NLSYNHLSGKIPPLYANKIY-AHDFIGNPGLC-----VDLDGLCR-----K 615  
QY 792 LAGSVAMG-----LLFSFVCLFGLILVGRMRKRRKKEALEMYAEGHNSGDRP 842  
Db 616 ITRSKNIGYVILITFLFGLVFWVGIVMFIAKCRKLRALKSTLAA----- 663  
QY 843 ANNTWMLTGVKEALISINLAFAFERPLRKLTF-----ADLQATNGFHNDLSIGSGGPDY 897  
Db 664 ---SKW-----RSFHKLHPSHEHIDCLDEKN-----VIGFGSSGRV 697  
QY 898 YKALIKGSSAVAIKKLIVHSGQDRE-----FMAEMETIGIKIRNLVPLIGYCK 947  
Db 698 YKVELRGGEVVAVKLKNKSVYKGDDEYSSDSLNRDVAFAEVEITLGRHKSIVRLWCCCS 757  
QY 948 VGERLIVNEVMKYGSLEVDYLOPKKGGVCLKSTRKTAIGSARGLAFLHNCSPHITH 1007  
Db 758 SGDCKLIVYEMRPGSLADVLHGDRKGGVLCWPERLIALDNAEGLSYLHHDVCPPIVH 817  
QY 1008 RDMSSVNLIDENLEAVSDFGMARL--MSAMDTLHLSVSTLAGTPGYVPEYYOSFRGCT 1065  
Db 818 RDVSSNILLDSYGAADFGIAKVGQMSGSKTPRAMSGIAGSCYIAPERYVTLRYNE 877  
QY 1066 KGDVYSYGVVLELLTGKRPDSDPDGNNLVGVV-KQHAKLISDVFPPELMKEDPALE 1124  
Db 878 KSDIYSFGVVLLELVGKQPTDS-ELGDKDMAMKVCCTALDKGLBVIDPKL--DLKFK 933  
QY 1125 IELLQHLKVAVACLDRAMRPTMVOVMAMFKEIQAGSGIDQSOTIRSIEDGG 1177  
Db 934 EELSKVYHIGLCTSPPLNRPSPMRKVIMLQEVSGAVPCSSPNTSKRSKXTGG 986

Completed: March 10, 2003, 18:24:52  
Time: 41.2258 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 17:58:34 ; Search time 25.6693 Seconds

(without alignments)  
1964.817 Million cell updates/sec

Title: US-09-823-394-2  
Perfect score: 6183  
Sequence: 1 MKTSSFFLSVTLFFPSF.....GSTLEMDMSIKEVDEGL 1196

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 188354 seqs, 42170167 residues

number of hits satisfying chosen parameters: 188354

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6183	100.0	1196	US-09-823-394-2	Sequence 2, Appl
2	1284.5	20.8	960	US-08-910-386A-18	Sequence 18, Appl
3	1087	17.6	1025	US-08-910-386A-7	Sequence 7, Appl
4	1055	17.1	1161	US-08-910-386A-20	Sequence 20, Appl
5	1009.5	16.3	996	US-08-910-386A-5	Sequence 5, Appl
6	937	15.2	854	US-09-754-853A-1111	Sequence 1111, Ap
7	937	15.2	854	US-09-754-853A-1113	Sequence 1113, Ap
8	937	15.2	854	US-09-754-853A-1115	Sequence 1115, Ap
9	937	15.2	877	US-09-754-853A-1110	Sequence 1110, Ap
10	937	15.2	877	US-09-754-853A-1112	Sequence 1112, Ap
11	937	15.2	854	US-09-754-853A-1114	Sequence 1114, Ap
12	933	15.1	854	US-09-754-853A-1108	Sequence 1098, Ap
13	933	15.1	854	US-09-754-853A-1101	Sequence 1101, Ap
14	933	15.1	854	US-09-754-853A-1103	Sequence 1103, Ap
15	933	15.1	854	US-09-754-853A-1105	Sequence 1105, Ap
16	933	15.1	854	US-09-754-853A-1107	Sequence 1107, Ap
17	933	15.1	877	US-09-754-853A-1109	Sequence 1109, Ap
18	933	15.1	877	US-09-754-853A-1100	Sequence 1100, Ap
19	933	15.1	877	US-09-754-853A-1102	Sequence 1102, Ap

20	933	15.1	877	9	US-09-754-853A-1104	Sequence 1104, Ap
21	933	15.1	877	9	US-09-754-853A-1106	Sequence 1106, Ap
22	932	15.1	854	9	US-09-754-853A-1109	Sequence 1109, Ap
23	932	15.1	877	9	US-09-754-853A-1108	Sequence 1108, Ap
24	822.5	13.3	894	9	US-09-754-853A-1099	Sequence 1099, Ap
25	822.5	13.3	894	9	US-09-754-853A-1116	Sequence 1116, Ap
26	822.5	13.3	894	9	US-09-754-853A-1117	Sequence 1117, Ap
27	821.5	13.3	894	9	US-09-754-853A-1119	Sequence 1119, Ap
28	817.5	13.2	894	9	US-09-754-853A-1118	Sequence 1118, Ap
29	763	12.3	813	8	US-08-910-386A-9	Sequence 9, Appl
30	656.5	10.6	612	8	US-08-910-386A-2	Sequence 2, Appl
31	629.5	10.2	830	10	US-09-772-134B-14	Sequence 14, Appl
32	617.5	10.0	749	10	US-09-828-113-27	Sequence 27, Appl
33	614.5	9.9	731	9	US-10-086-464-17	Sequence 17, Appl
34	604.5	9.8	647	9	US-10-086-464-2	Sequence 2, Appl
35	604.5	9.8	647	9	US-10-086-464-4	Sequence 4, Appl
36	604.5	9.8	721	9	US-10-086-464-5	Sequence 5, Appl
37	601	9.7	633	9	US-10-086-464-11	Sequence 11, Appl
38	595.5	9.6	674	9	US-10-086-464-14	Sequence 14, Appl
39	586	9.5	731	9	US-10-086-464-8	Sequence 8, Appl
40	581.5	9.4	355	10	US-09-862-027-26	Sequence 26, Appl
41	505	8.2	604	10	US-09-862-027-17	Sequence 17, Appl
42	499	8.1	645	10	US-09-862-027-15	Sequence 15, Appl
43	499	8.1	645	10	US-09-862-027-16	Sequence 16, Appl
44	482.5	7.8	339	10	US-09-838-955-8	Sequence 8, Appl
45	481	7.8	669	10	US-09-862-027-27	Sequence 27, Appl

#### ALIGNMENTS

RESULT 1	US-09-823-394-2	US-09-823-394-2
1	Sequence 2, Application US/09823394	
2	Publication No. US20030041344A1	
GENERAL INFORMATION:		
APPLICANT:	Chory, Joanne	
APPLICANT:	Jianming, Li	
APPLICANT:	Salk Institute for Biological Studies	
TITLE OF INVENTION:	RECEPTOR KINASE, BIN 1	
FILE REFERENCE:	SALKINS.012CPI	
CURRENT APPLICATION NUMBER:	US/09/823,394	
CURRENT FILING DATE:	2001-03-30	
PRIOR APPLICATION NUMBER:	08/881,706	
PRIOR FILING DATE:	1997-06-24	
NUMBER OF SEQ ID NOS:	2	
SOFTWARE:	FastSeq for Windows Version 4.0	
SEQ ID NO 2		
LENGTH:	1196	
TYPE:	PRT	
ORGANISM:	Arabidopsis	
US-09-823-394-2		
Query Match	100.0%; Score 6183; DB 9; Length 1196;	
Best Local Similarity	100.0%; Pred. No. 0;	
Matches 1196; Conservative	0; Mismatches	0; Indels
0; Gaps	0;	
QY	1 MKTSSFFLSVTLFFPSFSLSPQSLYRETHQISPDVY.PDKRLPDWSSNKK	60
DB	1 MKTSSFFLSVTLFFPSFSLSPQSLYRETHQISPDVY.PDKRLPDWSSNKK	60
QY	61 PCTPDGVTCRRDKTYSIDLSKPLNVGSAVSSLLSLTGLBSLFLSNHINGSVGFKC	120
DB	61 PCTPDGVTCRRDKTYSIDLSKPLNVGSAVSSLLSLTGLBSLFLSNHINGSVGFKC	120
QY	121 SASLTSIDLSNLSGCVTLTSLGSCGLKFLNVSSNTLPDPGRKVSGLKNSLFEVLDL	180
DB	121 SASLTSIDLSNLSGCVTLTSLGSCGLKFLNVSSNTLPDPGRKVSGLKNSLFEVLDL	180
QY	181 SANSIGANVGVWLSGCGELKRLAISGNKISGVDVSRCVNLEFLDVSSNNFSTGTFPF	240
DB	181 SANSIGANVGVWLSGCGELKRLAISGNKISGVDVSRCVNLEFLDVSSNNFSTGTFPF	240

```

QY 241 LGDCSALQHLIDISGNKLSGDFSRISTCTELKLNLSNQFVGP1PPULKSLQYLISLAE 300
Db 241 LGDCSALQHLIDISGNKLSGDFSRISTCTELKLNLSNQFVGP1PPULKSLQYLISLAE 300
QY 301 NKFTGPIPDPLSGACDTLTGLDISGNHFGAVPPFFGSCSLLESALSSNNNSGGLPMDT 360
Db 301 NKFTGPIPDPLSGACDTLTGLDISGNHFGAVPPFFGSCSLLESALSSNNNSGGLPMDT 360
QY 361 LLMORGLKVLDSLSPNEFSGELPESLTLNLSASLLTLTDLSSNNSGGLPMLCONPKXTLOE 420
Db 361 LLMORGLKVLDSLSPNEFSGELPESLTLNLSASLLTLTDLSSNNSGGLPMLCONPKXTLOE 420
QY 421 LYLQNGFTGKIPTLISNCELSVLSLSPNYLSGTLPSISGSLSKRLDKLMLNLEGEI 480
Db 421 LYLQNGFTGKIPTLISNCELSVLSLSPNYLSGTLPSISGSLSKRLDKLMLNLEGEI 480
QY 481 POELMYVKTLETLILDNDLTGELIPSGLSNCTNLMNLSISNNRLTGBIPKMLNLEGEI 540
Db 481 POELMYVKTLETLILDNDLTGELIPSGLSNCTNLMNLSISNNRLTGBIPKMLNLEGEI 540
QY 541 LKLSNNSFSGNIPDELDCRSLIWLDTNMLFNGTIPAMFKOSGKIAANFIAGKRYVYI 600
Db 541 LKLSNNSFSGNIPDELDCRSLIWLDTNMLFNGTIPAMFKOSGKIAANFIAGKRYVYI 600
QY 601 KNDGKKECHGAGNLLEFQGISFQNLSTRNPCNITSRVYGHTSPPTDNNKSMWFLD 660
Db 601 KNDGKKECHGAGNLLEFQGISFQNLSTRNPCNITSRVYGHTSPPTDNNKSMWFLD 660
QY 661 MSYNMLSGYIPKEIGSMPLYLTLNIGHNDISGSIPEDEVDLKGNTLIDLSNNKLDGRIPQ 720
Db 661 MSYNMLSGYIPKEIGSMPLYLTLNIGHNDISGSIPEDEVDLKGNTLIDLSNNKLDGRIPQ 720
QY 721 AMSALTMLTEIDLSNNLSGPIPEWGFETPPAKFLNPGLCYPLPCDPSNADGIAH 780
Db 721 AMSALTMLTEIDLSNNLSGPIPEWGFETPPAKFLNPGLCYPLPCDPSNADGIAH 780
QY 781 HORSGRPPASLAGSVANGLLFSPYCFGLILVGRKRRKREKAELEMYAEGHSGSD 840
Db 781 HORSGRPPASLAGSVANGLLFSPYCFGLILVGRKRRKREKAELEMYAEGHSGSD 840
QY 841 RTANNNTKLTGVKALSLINLAEPKPLKTLFADLLQATNGFHNDSLISGGFGDYKA 900
Db 841 RTANNNTKLTGVKALSLINLAEPKPLKTLFADLLQATNGFHNDSLISGGFGDYKA 900
QY 901 ILKDSAAVAIKKLHVSGQGRFPAEMETIGKIKRNLVPLGLGCKVGDRLNLEVMK 960
Db 901 ILKDSAAVAIKKLHVSGQGRFPAEMETIGKIKRNLVPLGLGCKVGDRLNLEVMK 960
QY 961 VGSLEVDYODPKKGVKCLKSTRKRIATGSARGLAFLHHNCSPIIHRDMKSSNVLDEN 1020
Db 961 VGSLEVDYODPKKGVKCLKSTRKRIATGSARGLAFLHHNCSPIIHRDMKSSNVLDEN 1020
QY 1021 LBARVSDFGMARLMSAMDTHLSVSTLACTPGYVPEYYSQFRCSTKGDVYSYGVLLLEL 1080
Db 1021 LBARVSDFGMARLMSAMDTHLSVSTLACTPGYVPEYYSQFRCSTKGDVYSYGVLLLEL 1080
QY 1081 TGRKRPDSDPGDNNLVGVWVOKHAKRLISDVDPBELMKEDPALREILLOHLVAVACLD 1140
Db 1081 TGRKRPDSDPGDNNLVGVWVOKHAKRLISDVDPBELMKEDPALREILLOHLVAVACLD 1140
QY 1141 RANRRPTMVOVMAFKEIOAGSGIDSGSTRSIEDGFSSTIEMVMSIKEVEGKL 1196
Db 1141 RANRRPTMVOVMAFKEIOAGSGIDSGSTRSIEDGFSSTIEMVMSIKEVEGKL 1196

```

RESULT 2  
US-08-910-386A-18

; Sequence 18, Application US/08910386A  
; Patent No. US20020092041A1

; GENERAL INFORMATION:  
; APPLICANT: Ronald, Pamela C.

; APPLICANT: Wang, Guo-Liang  
; APPLICANT: Song, Wen-Yuang

```

; APPLICANT: Hulbert, Scot
; APPLICANT: Richter, Todd
; TITLE OF INVENTION: Procedures and Materials for Confering
; TITLE OF INVENTION: Disease Resistance in Plants
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,386A
; FILING DATE: 13-AUG-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Baebian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 023070-058950US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 960 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-910-386A-18

```

Query Match 20.8%; Score 1284.5; DB 8; Length 960;

Best Local Similarity 32.1%; Pred. No. 9.3e-75; Indels 257; Gaps 35;

Matches 378; Conservative 168; Mismatches 373;

```

QY 10 SVTLFFPFSFSLFOASPSQ---SLYREIHQILSPFDVLPDKNL---LPDM--SSNKN 60
Db 4 SNTSLFFPFLVFLVLTIPSSQSNLSLRQAKTVLSKYAFVQSSVPSTLSNMNNSNYS 63
QY 61 PCTFDVTCRDK-VTIDLSKPLNVGSAVSSLSLTGLLESPLNSHNGSVSGFK 119
Db 64 ICSWGTICDDTSVTSIDIS-----NLNISG-- 90
QY 120 CSASLTSLDSRNSLSGPTTLTSLGSCGKLFELNVSSNTLDPKGVGGLK-LNSLEVL 178
Db 91 -----SLSPDHELR-----LRVLNLSNL--FSGNLSWEYREFVVLQVL 129
QY 179 DLANSISGANVGVWVLSDCGELKHLAISGNKISGDVDVSRVNLLEFLDVSSNNPSTGI 238
Db 130 DAYNNPFSGLPLPG-----VTQLVQLYKYLNFQGNFSGKI 164
QY 239 PF-LGDCSALQHLIDISGNKLSGDFSRISTCTELKLNLS-SNOFVGP1PPUL-PIKSPQ 294
Db 165 PLVSGSFNQLLEPLSLAGNDLHGP1PRELGNVTSLRWLQGYTNOFDEG1PPELGKLVNLY 224
QY 295 YLGLAENKFTGELIPDFLSGACDTLTGLDISGNHFGAVPPFFGSCSLLESALSSNNFSG 354
Db 225 HLDSSCNLTGSLPPPL-GNLMNLDTLFLQKQTLGVFPQGNLRLKSLDISVIELTG 283
QY 355 ELPMDTLLKMRGLKVLDSLSPNEFSGELPESLTLNLSASLLTLTDLSSNNSGGLPMLCONP 414
Db 284 EIPVD-LSGKEILLNLFINNLHGEIPGCIABL-PLEMLNLMRNKFTGS1PSKLGMMG 341
QY 415 KNTLOELYLQNGFTGKIPTLISNCELSVLSLSPNYLSGTLPSISGSLSKRLDKLMLN 474
Db 342 K-LIEIDLSNNRLTGLIPKSLCRGNLKITLLDNFLFEPPLDDDGQCRITSRVAMGN 399
QY 475 MLBEIIPQELMYVKTLETLILDNDLTGELIPSGLSNCTN-LMNLSISNNRLTGBIPKMWIG 533

```



```

Db 400 YLSSGIPFGPIYELBELSVLELONNYISGQLNNEKSSASSKLEGLNLSNNRLSGALPSAIG 459
Qy 534 RLENIATIKLSNNSFSGNIPDELDCRSLIWLIDNTLIFNGTIPAMAFKQSGKIAANFIA 593
Db 460 NYSGKMLVLTNGNFGSDIPEDIGRLKSLIKLDSRNPFSGTIP----- 503
Qy 594 GKRYVYIKNDMKKECHGAGNLEFQGRSEQLNRSLSTRNPNCRISRYGHTSPPTDNN 653
Db 504 -----PQIGNC 509
Qy 654 GSNMFLMSYNNLSGYIPKEIGSWPYLFIPLGHNDISGIPDEVDLACGLNLDLSNK 713
Db 510 LSLTYLDSQNGISGPIPVQIAQHILNTYINISWNPESLPELIGLMS----- 559
Qy 714 LDGIPQMSALTMLTETIDLSNNLSGPIPEMGOFEFPFAKLNPPGCGYLPBCDPS 773
Db 560 -----LTSADFSHNNLSGSIPEGTQYLYFNSTPTGNYLSG-----SDSTPS 602
Qy 774 NADGYAHORSHG-----RRPASLGSVAMGLIFSVCTFGILVGRNKRKRKEAEL 828
Db 603 NITSNSSEIGDSDSRKTKVPTTKTFAFGLF---C-SLIFVLAITKTRKSGK--- 654
Qy 829 EMTAEGHSGSDRTANNMTKLTGVKEALSINLAPEKFLKLTFA--DLQATNGFHND 886
Db 655 -----NSNMLKL-----AFQ-----KLFGSBDVLQC--LIDN 681
Qy 887 SLISGGGFGDYKAILKQSAVAIKKLHVSQGGDEFMAEMETIGIKHRLVYLLGYC 946
Db 682 NVIGRGAGLIVYGTMPNDHVAVKLGISKSHDGLSAELNTLGKLRHRYVLLAF 741
Qy 947 KYVDEBLVNEVWKYSLBDVLQDPKGGVYKLTSTRRIAGSARGLAFPHANCSPIIT 1006
Db 742 SNKEVNLVYETWNLNSLEVLHG--KNGGQLQWETRLKIALEAKGLSYLHHCSPWII 799
Qy 1007 HRMKSNNVLDENLEFARYSDFGMALMSAMDTHLSVSTLAGTPGYPEPYOSFSGCTK 1066
Db 800 HRVKSNNILNLSLEAHVADFLAKYFNNGTSEMSALNGSYVIAPEVAYLLKDEK 859
Qy 1067 GDVYSYGVVLELTLGKRPDSDPDGDN--NLVGVNKGAKLR--ISDVPPELMKEDP 1121
Db 860 SDVYSYGVVLELTLGKRPVGG--NFGEGMDIVQWAKETETKMSKGVYKIDERTL--KNV 915
Qy 1122 ALEIELQHLKVAVALDDBRAMRPTMYQVAMAFKE 1157
Db 916 AI-VEAMQVFVAMLCVEBYSIERPTMBSVQVOMLSQ 950

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,386A
FILING DATE: 13-AUG-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Baetian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 023070-058950US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1025 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-910-386A-7

Query Match 17.6%; Score 1087; DB 8; Length 1025;
Best Local Similarity 30.8%; Pred. No. 5e-62;
Matches 323; Conservative 140; Mismatches 371; Indels 216; Gaps 28;

Qy 231 SNFSTGI--PFLGDSALQHDHISGNKLSGDFSRASITCTELKLNINSQVGPPIPL 288
Db 88 SSNLS-GIISPGLNLSFLRELDGNDVLSGEIPELSRLSRLQLEISDVSIGSIPA- 145
Qy 289 PLKSLQYSLANKFTGEPFLSGACDTLTGLDLSGNHFYGAVPPEFG--SCSLIESIAL 347
Db 146 -----AIGACTKLTSLDLSHQLRGMIPREIGASLKLHLSLYL 183
Qy 348 SSNFSGLPMDTLKMGKLVLDSPNEFSGELPESTLNSASLTLDDSSNNFSGPIL 407
Db 184 YKNGLSGEIP-SALGNLTSLOEPDLSFRNLSGALPSLGLS-SLTMNIGQNNLSG-MI 240
Qy 408 PVLCON-----PKYTLQELV-----QNNFGTKIPPTSGSELY 443
Db 241 PMSINWLSLAFSVRENKLGMIPTNAFKTLHLEVIDMGTRFHOKTIAVSANASHLT 300
Qy 444 SLHSFNYLSGTPISLSLSKLDKLMNM----- 475
Db 301 VIQIGNLPFGIITGFRRLNLTLYLMLRNLPTREDDMGFISDLTNCISKIQTINLGE 360
Qy 476 --LGEIPEQELMYVT--LETLIDFNDLTGEIPSGNCTNLWISLNNRLTGEIPKI 532
Db 361 NNLGGLVLPNSFNLSSTSLALFLAKITGSIPODGNLIGLOHLYLCNNRFGSLPSL 420
Qy 533 GRLENLALIKLSNNSFSGNIPDELDCRSLIWLIDNTLIFNGTIPAMFKQSG---KIA 588
Db 421 GRKNTGLILAYENNLGSIPLAIGNLTENIILLGNNKSGWIPYTLNLTNLSLGLS 480
Qy 589 ANFLAGKRYVYIKNDMKKCHGAGNLEFQGRSQRLNRLSTRNPNCRISRYGHTSP 648
Db 481 TNNLSG-----IPSELFNIOQLSIMINVSNNLBSIPIQ 515
Qy 649 TFDNNGSMFLMSYNNLSGYIPKEIGSWPYLFIPLGHNDSGIPDEVDLACGLNLD 708
Db 516 EIGIKMLVEFHAESNRLSGKITNLTGDOQLRYLYLVQNNLSGSIPLSGQLKGLLETD 575
Qy 709 LSSNKLGRIPQMSALTMLTETIDLSNNLSGPIPEMGOFEFPFAKLNPPGCGYLP 764
Db 576 LSSNLSGQIPISLADITMHSINTLSFNSFVGEVPTIGAFSAASGISIQNAKLCGGIPD 635
Qy 765 YPLPRCDPSNADGYAHORSHGRRPASLAGSVAMGLIFSVCTFGILVGRNKRKRKEK 824
Db 636 LHLPRCP-----LLENRKGFPVLPISVSLAALAILSLYLI-----TWHRKTRK 682
Qy 825 EAELEMTAEGHSGSDRTANNMTKLTGVKEALSINLAPEKFLKLTFAFDLQATNGFH 884
Db 683 GABSRISMKEH-----PL--VYSQLVATQDFA 709
Qy 885 NDSLISGGGFGDYKAILKQSAVAIKKLHVSQGGDEFMAEMETIGIKHRLVYLLG 944

```

T 3  
 Sequence 7, Application US/08910386A  
 Patent No. US20020092041A1  
 GENERAL INFORMATION:  
 APPLICANT: Ronald, Pamela C.  
 APPLICANT: Wang, Guo-Liang  
 APPLICANT: Song, Wen-Yuang  
 APPLICANT: Hulbert, Scot  
 APPLICANT: Richter, Todd  
 TITLE OF INVENTION: Procedures and Materials for Confering  
 NUMBER OF SEQUENCES: 53  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Townsend and Townsend and Crew LLP  
 STREET: Two Embarcadero Center, Eighth Floor  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94111-3834  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30



QY 1035 SAMDTHLSVSTAGTGVYPPERYOSFRCTGADVSGVLLLELTGKRPTDSDPDGDN 1094  
 Db 1048 GTSETH-ATTGVAAGTGVYAPAYAMTCRVSDKADYVSGVLLLELISPKALD-PSFSSY 1105  
 QY 1095 ----NLVGV---VKOAKLRIS 1109  
 Db 1106 GNGFNIVAMACMLLRKAVLRSS 1127

RESULT 5  
 US-08-910-386A-5  
 ; Sequence 5, Application US/08910386A  
 ; Patent No. US20020092041A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ronald, Pamela C.  
 ; APPLICANT: Wang, Guo-Liang  
 ; APPLICANT: Song, Men-Yuang  
 ; APPLICANT: Hulbert, Scott  
 ; APPLICANT: Richter, Todd  
 ; TITLE OF INVENTION: Procedures and Materials for Confering  
 ; TITLE OF INVENTION: Disease Resistance in Plants  
 ; NUMBER OF SEQUENCES: 53  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: Townsend and Townsend and Crew LLP  
 ; STREET: Two Embarcadero Center, Eighth Floor  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94111-3834  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/910,386A  
 ; FILING DATE: 13-AUG-1997  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Baetian, Kevin L.  
 ; REGISTRATION NUMBER: 34,774  
 ; REFERENCE/DOCKET NUMBER: 023070-058950US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 576-0200  
 ; TELEFAX: (415) 576-0300  
 ; INFORMATION FOR SEQ ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 996 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-910-386A-5

Query Match 16.3%; Score 1009.5; DB 8; Length 996;  
 Best Local Similarity 30.5%; Pred. No. 4,66-57;  
 Matches 310; Conservative 148; Mismatches 391; Indels 167; Gaps 29;

QY 161 DPGKVGSGKLKINSLEVDLSANSI--SGANVVGWVLSDG-----CGELKH----- 204  
 Db 24 DDDGDAAG----DEALLSFKSSLLYGGGSLASWNISGHQHCWGVVGR-RHHRV 78  
 QY 205 --LAISGKIGSDVDVRCNVLEF--LDVSSNNFSTGIP-FLDCCSALOHLDISGKLS 258  
 Db 79 VKLRKRSNLTGIIISPS-IGNLSFKRTQLSNHNSGKIPQELSRSLQQLVNFNLS 137  
 QY 259 GDFSAISTCTELKLNINSNQFGRIPPL--PLKSLQYLSIANKKTGTEIPDLSGACD 316  
 Db 138 GEIPALGNLTSLVLEIHNNTLSGSISSIGKLTGLYNLALAEMLSGSIPTSF-QQLR 196  
 QY 317 TLVTGLDGNHFGVAVPFPFGSCSLLESIALSNNFSGELPMDTLLKMGKGLVDLSFNE 376  
 Db 197 RLSPFLSLAFNHLGSAIPDPINWISLTIFFEVVSNLTOTLPANAFNLPNLOQVPMYNN 256

QY 377 FSGELPESTLNLSA-SILLTDLSSNNFSGPILPNLCQNPXTLOEYL-----QNNG 427  
 Db 257 FHGPISASINASISIFITGL--NSFSGVVPPEI--GMRNLQRLLEFETLLEAETND 312  
 QY 428 FTGKIPTLSNCELYSLHSFNYSGLTIPSSIGSL-SKLRLDKMLNMLEBIPQELMY 486  
 Db 313 W--KFM TALNCSNLOEVLJAGCKFGGVLPDSVSNLSLSISIRDNKISGLSPRIDGN 370  
 QY 487 VKTLETLIDFNDLTGEIPSGLSNCTNLMWISLNNRLTGEIPKXIGRENLAILKLSN 546  
 Db 371 LVNLQYISLANNLSFOSLPSFSFKLKNLRRLTYDNNRLTGLPLTIGNLTQLTNMEVQFN 430  
 QY 547 SFGNTPDELGDCKSLIWLDTNLNFGTIPAMF-----KOSGLAANFIAGKRYVIK 601  
 Db 431 AFGTIPSTLGNLTGLFQNLGNHNNFTGQIPLEIFSPALSELIDVSHNNLG----- 483  
 QY 602 NDMKKECHGAGMLRFQGRISQRLRSTRPNCNITSRYVGHTSPYDNNGSMFLDM 661  
 Db 484 --SIPKEIGKKNIVEFHA----- 500  
 QY 662 SYNNLSGYTPKEIGSMPYFLINLGHNDISGIPDEVDLRGINTLDSNKLDRIPQA 721  
 Db 501 DSKLGEIPSTTIGECQLQHLFLQNNFNGSIPALITOLKGLDITLDSGNNLSGQIPWS 560  
 QY 722 MSALITMLTEIDLSNNLSGPIPEMGQFETPPAKFLNPNLGC---YPLPRCDPSNADG 777  
 Db 561 LGDMTLLHSNLNSFNSFHBGVPTNGVFANASEIYIQGNHICGAIPELHLPTCS----- 614  
 QY 778 YAHQSHGRRRPPLASGAYMGLBSFVCI FGLILVGRMRRRKKEALEMYABHGN 837  
 Db 615 ---LKSRRKRKHQIILLVWVICLVSTLAVFSLVYMLTCHK-RRKEVPATYSQGH-- 667  
 QY 838 SGDRYANNNTWKLTVKKAALSNLAFAEKPLKLPADLOATNGHNDLSLGSFGGV 897  
 Db 668 -----PM--ITYKQVKAIDGSSSHLGSQSPGSV 696  
 QY 898 YKAIL--KDG---SAVAIKKLIVHSGGDRFEMAEMETIGIKRHNLYVLLGYCYKGR 952  
 Db 697 YKGFPSQGEITSLVAVAVKLETPKALKSTABCEITLRNRHNVIVIVICSIDNR 756  
 QY 953 -----LVNENWVKYSGLEVLDPPKKGVK---LKSTRTKKAISGARGLAFHNCSPH 1004  
 Db 757 GNDPKAIVYDFENGSLEBMLPBTNDQAEQRHLTHQVTLILVACALEHILFHGPEP 816  
 QY 1005 IIRDMKSSNVLLDENLBARVSGFGKARLM---SAMDTHLSVSTLAGPGVPPERYOS 1060  
 Db 817 IYCHDKSSNVLLDPMVAHVGFGLARILVGGSSLMQOSTSSMGIRGTIGYAAPEYGVG 876  
 QY 1061 FRCTGADVSYGVVLLLELTGKRPTDSDPDGDNVLVGVVKOAKLRISDVPPEL 1116  
 Db 877 NTASTHGDIYSGLIVLETVTGMRPADSTFRITGLSLRQYVEGHLGRIMDVDRRL 932

RESULT 6  
 US-09-754-853A-1111  
 ; Sequence 1111, Application US/09754853A  
 ; Publication No. US20030005491A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hauge, Brian M.  
 ; APPLICANT: Parsons, Jeremy D.  
 ; APPLICANT: Wang, Ming Li  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Soybean Cyst Nematode Resistance  
 ; FILE REFERENCE: 38-10(15810)B  
 ; CURRENT APPLICATION NUMBER: US/09/754,853A  
 ; PRIOR FILING DATE: 2001-01-05  
 ; PRIOR APPLICATION NUMBER: US 60/174,880  
 ; NUMBER OF SEQ ID NOS: 1119  
 ; SEQ ID NO 1111  
 ; LENGTH: 854



Db 611 SEVSVLGKIRHNPVIALRAY-YLGRKGEKLLVPDYMKSGLASTFLHG---CGTETFDWP 666  
 Qy 982 TRKTAIGSARGLAFLHNNCSPIIHRDMKSSNVLLDENLEARYSDGMAIAMSAMDTHL 1041  
 Db 667 TRMKTAQDLARGFLCLH--SQENIIHGNTSSNVLLDENNTAKIADFGLSIMSTA-ANS 723  
 Qy 1042 SVSTLAGTPGVPPYVYQSFRCSTKGDVYSYGVVLELTGKREPTSDPDGNNLVGMVK 1101  
 Db 724 NVIATAGALGYRAPELSKTKKANTKTDIYSGLVILELTTRKSP--GVPMNGDLDPQVWA 781  
 Qy 1102 QHAKLR-ISDVPDELMKEDPALIEILLQHLKVAVACLDDBAMRPRTMVOYMAFKEIQ 1159  
 Db 782 SVVKEEMTNEVFADLMRDASTVGDDELNTLTKALHCVDPSPARPEVHQVLOGLEIR 840

## RESULT 8

US-09-754-853A-1115  
 ; Sequence 1115, Application US/09754853A  
 ; Publication No. US20030005491A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hauge, Brian M.  
 ; APPLICANT: Parnell, Laurence D.  
 ; APPLICANT: Parsons, Jeremy D.  
 ; APPLICANT: Wang, Ming Li  
 ; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
 ; FILE REFERENCE: 38-10(15810)B  
 ; CURRENT APPLICATION NUMBER: US/09/754,853A  
 ; PRIOR FILING DATE: 2001-01-05  
 ; PRIOR APPLICATION NUMBER: US 60/174,880  
 ; PRIORITY FILING DATE: 2000-01-07  
 ; NUMBER OF SEQ ID NOS: 1119  
 ; SEQ ID NO 1115  
 ; LENGTH: 854  
 ; TYPE: PRF  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: rhg1\_pi200499\_amplicon  
 ; US-09-754-853A-1115

Query Match 15.2%; Score 937; DB 9; Length 854;  
 Best Local Similarity 31.3%; Pred. No. 1,7e-52;  
 Matches 281; Conservative 113; Mismatches 279; Indels 226; Gaps 24;

Qy 290 LKSLQVLSAENKFTGEIPDLFSGACDTLTGLDLSGNHFGAVPPFGSCSLLESIALSS 349  
 Db 139 LQGLRKSLHDNDIGSIPSTL-GLLPNLRGVQLFNNRLTGSIPLSIGFCPLQSDLSN 197  
 Db 350 NNFSGELPMDTLTKMRGLKVLDSFNEFSGELPESLTNLSASILTLDSSNNFSGPILPN 409  
 Db 198 NLITGAIY-SLANSTKLVWLNSTFNSFSGPLASLTH-SFSLTFLSLQNNNLSSGLPNS 255  
 Qy 410 LCONPNK---TLQELYLQNNGFTGKIPPTLSNCSSELVSLHLSFNYSGLTIPESLSGLSL 466  
 Db 256 WGNNSKNGFFRQNLILDNPFQDVPAISGLRLNELISLSHNKFSGAIPEIGITLSRL 315  
 Qy 467 RDLKLMANLEBGIPELMTYVETLTLIDPDVLTCEIISGSLNCTNANWISLNNRLTG 526  
 Db 316 K-----TLDISNNAALNGNLPAATLSNLSLTLAAENNLIDN 351  
 Qy 527 EIPKWIIGRLNLAILKLSNNSFSGNIPDELGDCRSIIMDLNTNLFNGTIPAAFKQSGK 586  
 Db 352 QIPQSIGRLRNLSVILSNQFSGHIPSIANISSLRQDLDSLNNFSGEIPV----- 403  
 Qy 587 IAAPIAGRRYYIIRKDKKKECHAGNLIEFQGISRSEQLNLSLRPNCTITSRYVGHT 646  
 Db 404 ----- 403  
 Qy 647 SPTFNNNGSMFELMSYNNLSGYIPKEIGSMXYLFIPLNGHNDIGSIPDEVGLRGINT 706  
 Db 404 ---SPTDSQRSLNLFNYSNLSGSP-----PL----- 429  
 Qy 707 LDISNKLIDGRIPOAMSAITMLTEIDLSNNNLGPIPEWQGFETPPPAKFIANNPLGCGY- 765

Db 430 -----AKFNSSPFQNIOLCIS 448  
 Qy 766 PLPRC-----DSNADVAHQSHGRRPASLAGSVAMGLFSFYCIFLIWG 814  
 Db 449 PSTPCLSQAPSGQVIAPPEVYSKHHHRKLGTDKIIILVAGVLVLLIILCVLLFCL- 507  
 Qy 815 REMRKRKRKEAELEMYABEGHNSGD-----RTANNTNMKLITGVKEA---LSITLAAFE 865  
 Db 508 ---KRRSTK-----AGNQATEGPAATMRTEKVPVPAAGDVERGAGGKLVTFD 556  
 Qy 866 KPLRLKTFADILLQATNGRHNDLSIGSGFGVYKAILKDGSAVAIKLIHVSQGDREPM 925  
 Db 557 GPM-AFTADLLCAT-----AEMKSTYGVYKAILLEDGQVAVKRLREKITGKHREFE 610  
 Qy 926 AEMETIGIKIRHNPVIALRAY-YLGRKGEKLLVPDYMKSGLASTFLHG---CGTETFDWP 666  
 Db 611 SEVSVLGKIRHNPVIALRAY-YLGRKGEKLLVPDYMKSGLASTFLHG---CGTETFDWP 666  
 Qy 982 TRKTAIGSARGLAFLHNNCSPIIHRDMKSSNVLLDENLEARYSDGMAIAMSAMDTHL 1041  
 Db 667 TRMKTAQDLARGFLCLH--SQENIIHGNTSSNVLLDENNTAKIADFGLSIMSTA-ANS 723  
 Qy 1042 SVSTLAGTPGVPPYVYQSFRCSTKGDVYSYGVVLELTGKREPTSDPDGNNLVGMVK 1101  
 Db 724 NVIATAGALGYRAPELSKTKKANTKTDIYSGLVILELTTRKSP--GVPMNGDLDPQVWA 781  
 Qy 1102 QHAKLR-ISDVPDELMKEDPALIEILLQHLKVAVACLDDBAMRPRTMVOYMAFKEIQ 1159  
 Db 782 SVVKEEMTNEVFADLMRDASTVGDDELNTLTKALHCVDPSPARPEVHQVLOGLEIR 840

## RESULT 9

US-09-754-853A-1110  
 ; Sequence 1110, Application US/09754853A  
 ; Publication No. US20030005491A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hauge, Brian M.  
 ; APPLICANT: Parnell, Laurence D.  
 ; APPLICANT: Parsons, Jeremy D.  
 ; APPLICANT: Wang, Ming Li  
 ; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
 ; FILE REFERENCE: 38-10(15810)B  
 ; CURRENT APPLICATION NUMBER: US/09/754,853A  
 ; PRIOR FILING DATE: 2001-01-05  
 ; PRIOR APPLICATION NUMBER: US 60/174,880  
 ; PRIORITY FILING DATE: 2000-01-07  
 ; NUMBER OF SEQ ID NOS: 1119  
 ; SEQ ID NO 1110  
 ; LENGTH: 877  
 ; TYPE: PRF  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: rhg1\_noir\_amplicon  
 ; US-09-754-853A-1110

Query Match 15.2%; Score 937; DB 9; Length 877;  
 Best Local Similarity 31.3%; Pred. No. 1,8e-52;  
 Matches 281; Conservative 113; Mismatches 279; Indels 226; Gaps 24;

Qy 290 LKSLQVLSAENKFTGEIPDLFSGACDTLTGLDLSGNHFGAVPPFGSCSLLESIALSS 349  
 Db 162 LQGLRKSLHDNDIGSIPSTL-GLLPNLRGVQLFNNRLTGSIPLSIGFCPLQSDLSN 220  
 Qy 350 NNFSGELPMDTLTKMRGLKVLDSFNEFSGELPESLTNLSASILTLDSSNNFSGPILPN 409  
 Db 221 NLITGAIY-SLANSTKLVWLNSTFNSFSGPLASLTH-SFSLTFLSLQNNNLSSGLPNS 278  
 Qy 410 LCONPNK---TLQELYLQNNGFTGKIPPTLSNCSSELVSLHLSFNYSGLTIPESLSGLSL 466  
 Db 279 WGNNSKNGFFRQNLILDNPFQDVPAISGLRLNELISLSHNKFSGAIPEIGITLSRL 338

```

QY 467 RDLKMLNMLEGEIPEQELMYKTLLETLILDPNDLGLGELPSGLNCTNLTNMTSLSNRRLTG 526
Db 339 K-----TLDI SNNA LGNLPATLSNLSLTLLANAENNLIDN 374
QY 527 EIPKMGILENLAIAIKLNNNSFSGNI PDELGCGRSLIMDLINTLNFNGITPAAMEKOSGK 586
Db 375 QIPQSLGRNLRSVILIRNPSGHIPISSIANISSLRQDLDSLNFSGEIIPV----- 426
QY 587 IANFIACKRYVYIINKDKCKECHAGNLLFEGQIRSEQLNRLSTRNFCNITSRYGGHT 646
Db 427 ----- 426
QY 647 SPTFNNSNMFLDMSYMLSGYIPKEIGSMPLYFLIMLGHNDISGSIPDEVGLRGINI 706
Db 427 -SFDQSRLMNFVSNYSLGSVP-----PLL----- 452
QY 707 LDLSNKKLDGRIPQMSALTMLTEIDLNNNLISGPIPEMGQETEPPEPAKFLNPGLCGY 765
Db 453 -----AKKFVSSSFVGNILQCGYS 471
QY 766 PLPRC-----DPSNADGYAHQHSRGRRRPASLAGSVAMGLIFSVCLFGLILVG 814
Db 472 PPTPCISQAPSGVIAPPEVSGKHHHRKLTSTDIIILVAGVLLVLLILCCVLLFCLT- 530
QY 815 REMRRRRRKKKEALEYVYAGHNSGD-----RTANNTMKLTGYKEA--LSINLAPE 865
Db 531 ---RRKSTK-----AGNGQATEGRATMTEKGVPPVAAGDVEAAGEAGKGLVHD 579
QY 866 KPLRLKLPADLLQATGPNHDSILISGCGFDVYKAILKDSAAVAIKKLIVHSQGDREEM 925
Db 580 GPM-AFTADLLCAT-----AEIMGSKTYGYVYKAILLEDSQAVAVRLKEKTIKGRHEPE 633
QY 926 AEMETIGKIKHNLVPLLGYCKVG--DERLLVNEVWKYSGLEVDLPCKKGYK--LKLS 981
Db 634 SEVSVLGKIRHPVVALRAY-YLGPKEGKLLVADYMSKSLASFLLG--GGTEPIDWP 689
QY 982 TRRKIAIGSARGIAFLHNHNSPHIIRDKMSSVLLIDENLEAVSPFGMARLMSANDTHL 1041
Db 690 TRRKIAQDLARGLFCLH--SQENIIRHGNLTSSNVLLIDENNAKIADLGLRLMSTA-ANS 746
QY 1042 SVSTLAGTPGYVPEYXQSPFCSTKGDVYSYGVLLLELTGKRPDPSPDGDNNLVGWVK 1101
Db 747 NYATGALGYRAPBELSKLKKANTKDIYSLGYILLELIRKSP--GVPNNGLDLPQWVA 804
QY 1102 QHAKLR-ISDVPDELMKEDPALEIELLOHLKVAVACLDRAMRPRPTVYVAMFEIQ 1159
Db 805 SVYKEWTVNEVPADLMRDASTVGDELLNTLKIALHCVDSPSARPEVHQVLQDLBEIR 863
QY 1159 -----
Db 863 -----
QY 1159 US-09-754-853A-1112
; Sequence 1112, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hange, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; PRIOR FILING DATE: 2001-01-05
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 1112
; LENGTH: 877
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: rhgl_lee_amplicon
US-09-754-853A-1112

```

[illegible]

/ APPLICANT: Parsons, Jeremy D.  
 / APPLICANT: Wang, Ming Li  
 / TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
 / FILE REFERENCE: 38-10(15810) B  
 / CURRENT APPLICATION NUMBER: US/09/754,853A  
 / PRIOR FILING DATE: 2001-01-05  
 / PRIOR APPLICATION NUMBER: US 60/174,880  
 / NUMBER OF SEQ ID NOS: 1119  
 / SEQ ID NO 1114  
 / LENGTH: 877  
 / TYPE: PRT  
 / ORGANISM: Glycine max  
 / FEATURE:  
 / OTHER INFORMATION: Clone ID: rhg1\_p1200499\_amplicon  
 / US-09-754-853A-1114

Query Match 15.2%; Score 937; DB 9; Length 877;  
 Best Local Similarity 31.3%; Pred. No. 1.8e-52;

Matches 281; Conservative 113; Mismatches 279; Indels 226; Gaps 24;

QY 290 LKSIQVSLANKKFTGEPDPLSGACDTLTGLDSGNHFGYAVPFFGSCSLLESALSS 349  
 Db 162 LQGRKSLHNDQIGSIPSTL-GILPNLRGVOLFNNRLTGISIPLSIGFCPLQSLDLSN 220  
 QY 350 NNFSGELPMDTLKRGKLVLDLSENFSGELPESLTNLSASLTTLDSNNFSGILPN 409  
 Db 221 NLTLGALPY-SLANSTKLWNLNLSFNSGELPASLTH-SFSLTFLSLQNNMLSGSLPNS 278  
 QY 410 LCONPKR--TLQELYLQNNGFTGKIPTLSNCELSVLSHLSFNYSGLTIPSLGSLSL 466  
 Db 279 WGSNSKNQGFRLQNLIDNFFTGVDVPSLSGLRLNEISLSHNKSGALPNEIGTSLSL 338  
 QY 467 RDLKLMNLMBEIPQELMYKLTETLIDPNDLTGEIPSGLSNCTNLNWSISNNRLTG 526  
 Db 339 K-----TLDISNNALNGNLPTLSNLSLTLLAENMLDN 374  
 QY 527 EIPKWLGRLENLAILKLSNNSFSGNIPDELGDCRSILWLDLNTNLFNGTIPAMFKQSGK 586  
 Db 375 QIPQSLGRLENLAILKLSNNSFSGNIPDELGDCRSILWLDLNTNLFNGTIPAMFKQSGK 586  
 QY 587 IAAPIAGKRYVYIKNDGMKKECHAGNLLFQGISBQNLRLSTRNPNCTTSRYVGHT 646  
 Db 427 ----- 426  
 QY 647 SPTFDNNGSNMFLDMSYNMLSGYIPKEISMPYLITLNGHNDISGIPDEVGLRGLNI 706  
 Db 427 --SFDQSRSLNLFVSYNSLSGVP--PLL----- 452  
 QY 707 LDISNKLDRIPQAMSAITMLTEIDLSNNNSGPIPEMGQETEPPEPAKFLNPGLCGY- 765  
 Db 453 ----- 452  
 QY 766 PLPRC-----DPSNADGYAHORSHGRPASLAGVAMGLLFSFVCIPTGLIIVG 814  
 Db 472 PSTPCLSQSPGSGVLAPEPVSKEHHHRKLTSTKDIILIVAGVLLVYLLICCVLLFCIL- 530  
 QY 815 REMRKRRKKEAELEMYAECHGNSGD-----RTANNTNMKLTGVKXA--LSLNLAAFE 865  
 Db 531 ---RKRSTSK-----AGNGQATEGPAATMRTKEGVPPVPAADVBAAGEAGKLVHFD 579  
 QY 866 KPLRLKLPDILLQATNGRHNDLSIGSGFGDYVYAILKDGSAVAIKKLIHVSQGDREPM 925  
 Db 580 GPM-AFTADDLICAT-----AEIMGKSTYGVYKAILLEDGQVAVKRLRKRTKGRHEFE 633  
 QY 926 AEMETIGIKIRKNLVPILGYCKVG--DERLLINVMKYGSLSEDTLQPPKGGYK--LKLS 981  
 Db 634 SEVSIVLGRIRHPNVALALAY-YLQPKGKLLVPDYMSKGSILASFLHG--GGTETPIIDWP 689  
 QY 982 TRRKAIASAGLAFLHNNCSPHIIRDMKSSNVLDELNTLFAVSDPGMARLMSANDTHL 1041  
 Db 690 TRMKIAQDLARGLFCLH--SQENIIHGNTLSSNVLDBENNAKIDAGLSRLMSTA-ANS 746

QY 1042 SVSTIAGTGVVPEVYQSFRCSTKGDVYGVVLLTGTGKPTDSPDGGDNLLYGMWK 1101  
 Db 747 NVIATVAGLGYRABELSKTKKANKTKTDIYSLGVLLELTRKSP--GVPMNGDLQGMWA 804  
 QY 1102 QHAKLR-1SDVPELMEKEDPALEIELLQTLKAVACLDPRAMRRPTMVGWAMFEIQ 1159  
 Db 805 SVVKEEWINEVFDADLNKDSIVYGDDELNTLKLALCHVPSPSARPEVHQVQLLEIR 863

## RESULT 12

US-09-754-853A-1098  
 / Sequence 1098, Application US/09754853A  
 / Publication No. US00030005491A1  
 / GENERAL INFORMATION:

/ APPLICANT: Hauge, Brian M.  
 / APPLICANT: Patneil, Laurence D.  
 / APPLICANT: Parsons, Jeremy D.  
 / APPLICANT: Wang, Ming Li  
 / TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
 / FILE REFERENCE: 38-10(15810) B  
 / CURRENT APPLICATION NUMBER: US/09/754,853A  
 / PRIOR FILING DATE: 2001-01-05  
 / PRIOR APPLICATION NUMBER: US 60/174,880  
 / NUMBER OF SEQ ID NOS: 1119  
 / SEQ ID NO 1098  
 / LENGTH: 854  
 / TYPE: PRT  
 / ORGANISM: Glycine max  
 / FEATURE:  
 / OTHER INFORMATION: Clone ID: 240017\_region\_G3  
 / US-09-754-853A-1098

Query Match 15.1%; Score 933; DB 9; Length 854;  
 Best Local Similarity 31.4%; Pred. No. 3.1e-52;

Matches 282; Conservative 114; Mismatches 277; Indels 226; Gaps 24;

QY 290 LKSIQVSLANKKFTGEPDPLSGACDTLTGLDSGNHFGYAVPFFGSCSLLESALSS 349  
 Db 139 LQGRKSLHNDQIGSIPSTL-GILPNLRGVOLFNNRLTGISIPLSIGFCPLQSLDLSN 197  
 QY 350 NNFSGELPMDTLKRGKLVLDLSENFSGELPESLTNLSASLTTLDSNNFSGILPN 409  
 Db 198 NLTLGALPY-SLANSTKLWNLNLSFNSGELPASLTH-SFSLTFLSLQNNMLSGSLPNS 255  
 QY 410 LCONPKR--TLQELYLQNNGFTGKIPTLSNCELSVLSHLSFNYSGLTIPSLGSLSL 466  
 Db 256 WGSNSKNQGFRLQNLIDNFFTGVDVPSLSGLRLNEISLSHNKSGALPNEIGTSLSL 315  
 QY 467 RDLKLMNLMBEIPQELMYKLTETLIDPNDLTGEIPSGLSNCTNLNWSISNNRLTG 526  
 Db 316 K-----TLDISNNALNGNLPTLSNLSLTLLAENMLDN 351  
 QY 527 EIPKWLGRLENLAILKLSNNSFSGNIPDELGDCRSILWLDLNTNLFNGTIPAMFKQSGK 586  
 Db 352 QIPQSLGRLENLAILKLSNNSFSGNIPDELGDCRSILWLDLNTNLFNGTIPAMFKQSGK 586  
 QY 587 IAAPIAGKRYVYIKNDGMKKECHAGNLLFQGISBQNLRLSTRNPNCTTSRYVGHT 646  
 Db 404 ----- 403  
 QY 647 SPTFDNNGSNMFLDMSYNMLSGYIPKEISMPYLITLNGHNDISGIPDEVGLRGLNI 706  
 Db 404 --SFDQSRSLNLFVSYNSLSGVP--PLL----- 429  
 QY 707 LDISNKLDRIPQAMSAITMLTEIDLSNNNSGPIPEMGQETEPPEPAKFLNPGLCGY- 765  
 Db 430 ----- 429  
 QY 766 PLPRC-----DPSNADGYAHORSHGRPASLAGVAMGLLFSFVCIPTGLIIVG 814

Db 449 PSTCLSQAPSGVIAPPEVSKHHHRKLTSTKOIILIVAGVLLVLLIILCCVLLFCL- 507  
 Qy 815 REMKRRRKEAELEMYAEGHNSGD-----RTANNTMVKLTGVKEA---LSINLAFE 865  
 Db 508 ---RKSTSK-----AGNGATEGRAATMTEKGVPPVAGGVVEAGGAGGLVHFD 556  
 Qy 866 KPLKLTLPADLLQATNGFHNDSLIGSGGFDVYKAILKDGSAVAIKLIVHSGQGREFM 925  
 Db 557 GPM-AFTADLLCAT-----AEIMGSTYGVYKAILLEDGSQAVARLREKIKGHRFE 610  
 Qy 926 AEMETIGIKHRNIVPLIGYCKVG--DERLLVNEVMKYSLEDVLDPPKGGVK--LKLS 981  
 Db 611 SEVSVLGKIRHPNVLLARAY-YLGPKEKLLVPDYMSKSLASFLHG--GGETFTDMP 666  
 Qy 982 TRKRIASARGAFLHNHNSPHIIRDMKSNVLLDENILEARVSDPGMARLMSAMTHL 1041  
 Db 667 TRKRIADLARGLFCLH--SQENIIGHNLTSNVLLDENITNAKADFGSLRMSA-ANS 723  
 Qy 1042 SVSTLAGTPGVPEYVYOSFRCSTKGDVYSYGVVLELLTGKRPDTPDPCDNNLVGVK 1101  
 Db 724 NVIATAGALGRAPBELSKLKANTKTIDYSLGVLLELLTRKSPGVSMNGLD--LPQWVA 781  
 Qy 1102 QHAKLR-ISDVPELMKEDEPALIEILLQHLKVAACLDPRARPRPTMVMANFKEIQ 1159  
 Db 782 SVYKEWNTNEVPADLMRDASTVGDDELNTLKALHCVDPSPARPEVHQVLOQLEIR 840

RESULT 13  
 US-09-754-853A-1101  
 ; Sequence 1101, Application US/09754853A  
 ; Publication No. US20030005491A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hauge, Brian M.  
 ; APPLICANT: Parsons, Jeremy D.  
 ; APPLICANT: Wang, Ming Li  
 ; APPLICANT: Parsons, Jeremy D.  
 ; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
 ; FILE REFERENCE: 38-10(15810)B  
 ; CURRENT APPLICATION NUMBER: US/09/754, 853A  
 ; PRIOR FILING DATE: 2001-01-05  
 ; PRIOR APPLICATION NUMBER: US 60/174,880  
 ; PRIOR FILING DATE: 2000-01-07  
 ; NUMBER OF SEQ ID NOS: 1119  
 ; SEQ ID NO 1101  
 ; LENGTH: 854  
 ; TYPE: PRT  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: rhg1\_A3244\_amplicon  
 US-09-754-853A-1101

Query Match 15.1%; Score 933; DB 9; Length 854;  
 Best Local Similarity 31.4%; Pred. No. 3,1e-52;  
 Matches 282; Conservative 114; Mismatches 277; Indels 226; Gaps 24;  
 Qy 290 LKSLQYLSAEKFTGEIDPLSGACDTLTGLDSCNHPYGVAPPPFGSCSLLESIALS 349  
 Db 139 LOGRLKLSLHDQIGSISTLTGLIPNRGVOLFNRKLRGSLPISLGCPGLQGLDLSN 197  
 Qy 350 NNPSGELPMDTLKMGKLVLDLSENFSGELPESITNLASILLTDLSSNNFSGPILPN 409  
 Db 198 NLTGAIPIY-SLANSRKLWLNLSFNSFGPLPASLTH-SFSLTFSLONNMISGLSPNS 255  
 Qy 410 LCONPYN--TLQELYLONNGFGKIPPTLSNGSELVSLHSFNYSAGTIPSLGSLKL 466  
 Db 256 WGNSSKNGFRLQNLILDHNFEGDVPASISGSLREINLSLSNKSAGALPHEIGLSTL 315  
 Qy 467 RDLKMLNMLEGELPQELMVYKLETLILDNDLTGEIPSGLSNCTNLNWLISLNNRLTG 526  
 Db 316 K-----TLDISNNMLNGLPPTLSNLSLTLNENMLLDN 351  
 Qy 527 EIPKWIQRLNLAALKLSNNSFGSNIPDELGDCLSLIWLDTNLTNGTIPAMFQSGK 586

Db 352 QIPQSLGRRLNLSVLLSNNGSGHIPSISIANISSRLQDLISLNNFSGEIPV----- 403  
 Qy 587 IANFTAGKRYIYIKNDGKKECHAGANLEFQGRSEQNLNLSFRNCNITRSVYGGHT 646  
 Db 404 ----- 403  
 Qy 647 SPFDNNNGSMFLDMSYNNLSGYIPKEIGSMPEYFLINLGHNDIGSLPDEVGLRLNI 706  
 Db 404 --SPDSGRSLNLFNVSYNLSGSPV-----PLL----- 429  
 Qy 707 LDLSNKLGRIPQAMSLTMLTEIDLNNNLSGPTPEMGOFETPPAKELNPPGLCGY- 765  
 Db 430 -----AKFSSSFVGNIQCGIS 448  
 Qy 766 PLPRC-----DPSNADGYAHORSHGRPASLAGSVAMGLIFSVCI FGLILVG 814  
 Db 449 PSTCLSQAPSGVIAPPEVSKHHHRKLTSTKOIILIVAGVLLVLLIILCCVLLFCL- 507  
 Qy 815 REMKRRRKEAELEMYAEGHNSGD-----RTANNTMVKLTGVKEA---LSINLAFE 865  
 Db 508 ---RKSTSK-----AGNGATEGRAATMTEKGVPPVAGGVVEAGGAGGLVHFD 556  
 Qy 866 KPLKLTLPADLLQATNGFHNDSLIGSGGFDVYKAILKDGSAVAIKLIVHSGQGREFM 925  
 Db 557 GPM-AFTADLLCAT-----AEIMGSTYGVYKAILLEDGSQAVARLREKIKGHRFE 610  
 Qy 926 AEMETIGIKHRNIVPLIGYCKVG--DERLLVNEVMKYSLEDVLDPPKGGVK--LKLS 981  
 Db 611 SEVSVLGKIRHPNVLLARAY-YLGPKEKLLVPDYMSKSLASFLHG--GGETFTDMP 666  
 Qy 982 TRKRIASARGAFLHNHNSPHIIRDMKSNVLLDENILEARVSDPGMARLMSAMTHL 1041  
 Db 667 TRKRIADLARGLFCLH--SQENIIGHNLTSNVLLDENITNAKADFGSLRMSA-ANS 723  
 Qy 1042 SVSTLAGTPGVPEYVYOSFRCSTKGDVYSYGVVLELLTGKRPDTPDPCDNNLVGVK 1101  
 Db 724 NVIATAGALGRAPBELSKLKANTKTIDYSLGVLLELLTRKSPGVSMNGLD--LPQWVA 781  
 Qy 1102 QHAKLR-ISDVPELMKEDEPALIEILLQHLKVAACLDPRARPRPTMVMANFKEIQ 1159  
 Db 782 SVYKEWNTNEVPADLMRDASTVGDDELNTLKALHCVDPSPARPEVHQVLOQLEIR 840

RESULT 14  
 US-09-754-853A-1103  
 ; Sequence 1103, Application US/09754853A  
 ; Publication No. US20030005491A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hauge, Brian M.  
 ; APPLICANT: Parsons, Jeremy D.  
 ; APPLICANT: Wang, Ming Li  
 ; APPLICANT: Parsons, Jeremy D.  
 ; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
 ; FILE REFERENCE: 38-10(15810)B  
 ; CURRENT APPLICATION NUMBER: US/09/754, 853A  
 ; PRIOR FILING DATE: 2001-01-05  
 ; PRIOR APPLICATION NUMBER: US 60/174,880  
 ; PRIOR FILING DATE: 2000-01-07  
 ; NUMBER OF SEQ ID NOS: 1119  
 ; SEQ ID NO 1103  
 ; LENGTH: 854  
 ; TYPE: PRT  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: rhg1\_peeking\_amplicon  
 US-09-754-853A-1103

Query Match 15.1%; Score 933; DB 9; Length 854;  
 Best Local Similarity 31.4%; Pred. No. 3,1e-52;  
 Matches 282; Conservative 114; Mismatches 277; Indels 226; Gaps 24;  
 Qy 290 LKSLQYLSAEKFTGEIDPLSGACDTLTGLDSCNHPYGVAPPPFGSCSLLESIALS 349  
 Db 139 LOGRLKLSLHDQIGSISTLTGLIPNRGVOLFNRKLRGSLPISLGCPGLQGLDLSN 197  
 Qy 350 NNPSGELPMDTLKMGKLVLDLSENFSGELPESITNLASILLTDLSSNNFSGPILPN 409  
 Db 198 NLTGAIPIY-SLANSRKLWLNLSFNSFGPLPASLTH-SFSLTFSLONNMISGLSPNS 255  
 Qy 410 LCONPYN--TLQELYLONNGFGKIPPTLSNGSELVSLHSFNYSAGTIPSLGSLKL 466  
 Db 256 WGNSSKNGFRLQNLILDHNFEGDVPASISGSLREINLSLSNKSAGALPHEIGLSTL 315  
 Qy 467 RDLKMLNMLEGELPQELMVYKLETLILDNDLTGEIPSGLSNCTNLNWLISLNNRLTG 526  
 Db 316 K-----TLDISNNMLNGLPPTLSNLSLTLNENMLLDN 351  
 Qy 527 EIPKWIQRLNLAALKLSNNSFGSNIPDELGDCLSLIWLDTNLTNGTIPAMFQSGK 586



```

QY 290 LKSLQVLSAENKFTGEIPDLISGACDTLTGLDLSGNHYGAVPPFFGSCSLLESIALSS 349
DB 139 LOGLRKLSLHDNIGSISIPSTL-GLLPNLRGVQLFNNRLTGSIPLSIGFPCPLQSLDLSN 197
QY 350 NNFSGELPMDTLTKMRGLKVLDSLSEFSGELPSISLTLNLASLTLTLSSNNNSGPIILN 409
DB 198 NLITGAIPIY-SLANSTKLYWNLISFNSFGPLPASLTH-SFSLTFLSLONNNLSGSLPNS 255
QY 410 LCONPKN---TLQELYLONNGFTGKIPTLSNCSLVSLSLHSPNYLSGTPSSLSISLKL 466
DB 256 WGSNSKNPFRLQNLILDNHFTGDVPASLSLRELNLISLSHNKFSGALPNIIGTSLRL 315
QY 467 RDLKMLNLMBEIPQELMYVLTLETLLDPNDLTGEIPSGLSNCTNLNMTSLSNRLTG 526
DB 316 K-----TLDISNNALNGMLPATLSLTLNLASLTLTLNANNLILDN 351
QY 527 EIPKWIENLAILKLSNNSFGNIPDELDCSLIWLIDNTNLNLTNGTTPAMFKOSK 586
DB 352 QIPQSLGRNLNLSVILSRNQSCHIPSISLANISLRLQDLISLNNFSGEIPV----- 403
QY 587 IANFIAGKRYVYIKNDGMKKECHGAGNLLEFGQIRSEQLNRLSTRNPNCTISRYVGH 646
DB 404 ----- 403
QY 647 SPFDNNGSMFLLDSYMLSGYIPKEIGSMPLYLTLNLGNDISGSDPEVDGLRLNT 706
DB 404 --SPDSQSLNLFNVSYNSLSGSP-----PLL----- 429
QY 707 LDISSNKLDGRIPQMSALTMLTEILDSNNNLGSPIPMGQFETPPAKFLNPGICGY- 765
DB 430 -----AKFNSSSFVGNILQCGYS 448
QY 766 PLPRC-----DPSNADGYAHHORSHGRRPASLAGSVAMGLLFSVCIFGLIIVG 814
DB 449 PSTPCLSQAPSGVIAPRPPEVSKHHHRKLSKTDILIIYAGVLLVLIILCCVLLFCLL- 507
QY 815 REMKRRKKEAELEMYABEGHNSGD-----RTANNWKLTGVXEA--LSINLAPE 865
DB 508 --RKRSTSK-----AGNGQATEGRAATWRTKGVPPVAGDVEAGGEGKLVHFD 556
QY 866 KPLRKLTPADLLQATNGFHNDSLIGSGFGDYKAILKOGSAVAIKLILHVSQGDREEM 925
DB 557 GPM-AFTADULLCAT-----AELMGKSTYGTIVYKALIEBGSQVAVRLEKIKGHRFE 610
QY 926 AEMETIGIKIRNLVPLGCKVG--DERLLVNEVWKYGSLEVDVQDRKGGVK--LKLS 981
DB 611 SEVSVLGKIRHPVNLALRAY-YLGPKEKLLVPDYMSKSLASFHGG--GGETFTIDWP 666
QY 982 TRRKIIAGSARGLAFLHNCSPHIIRDMKSSNVLLDENLEARVSDFGMARLMSAMDTL 1041
DB 667 TRMKIIADLARGLFCLH--SQENIIGHNLTSSNVLLDENTYAKIADFGSLRIMSTA-ANS 723
QY 1042 SVSTLAGTGVVPEVYOSFRCTGKYVSYGVVLELLTGKRPDSDPFGDNMLVGMVK 1101
DB 724 NVIATGALGYRAPBELSKLKANTKTIDISLGVILLETTRKSGVSNMGD--LPQWVA 781
QY 1102 QHAKLR-ISDVPDEIMKEDPALIEILLQKVAACLDDBAMRRPTMVMAMFKETIQ 1159
DB 782 SVYKEEMTEVEVDADLMDASTVGBELNLTALHLCVDPSPSAKPEVHVOQLQLEIR 840

```

```

; CURRENT APPLICATION NUMBER: US/09/754, 853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174, 880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 1105
; LENGTH: 854
; TYPE: PR
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: rhg1_toyosuzu_amllicon
US-09-754-853A-1105

Query Match      15.1%; Score 933; DB 9; Length 854;
Best Local Similarity 31.4%; Pred.No. 3.1e-52;
Matches 282; Conservative 114; Mismatches 277; Indels 226; Gaps 24;

QY 290 LKSLQVLSAENKFTGEIPDLISGACDTLTGLDLSGNHYGAVPPFFGSCSLLESIALSS 349
DB 139 LOGLRKLSLHDNIGSISIPSTL-GLLPNLRGVQLFNNRLTGSIPLSIGFPCPLQSLDLSN 197
QY 350 NNFSGELPMDTLTKMRGLKVLDSLSEFSGELPSISLTLNLASLTLTLSSNNNSGPIILN 409
DB 198 NLITGAIPIY-SLANSTKLYWNLISFNSFGPLPASLTH-SFSLTFLSLONNNLSGSLPNS 255
QY 410 LCONPKN---TLQELYLONNGFTGKIPTLSNCSLVSLSLHSPNYLSGTPSSLSISLKL 466
DB 256 WGSNSKNPFRLQNLILDNHFTGDVPASLSLRELNLISLSHNKFSGALPNIIGTSLRL 315
QY 467 RDLKMLNLMBEIPQELMYVLTLETLLDPNDLTGEIPSGLSNCTNLNMTSLSNRLTG 526
DB 316 K-----TLDISNNALNGMLPATLSLTLNLASLTLTLNANNLILDN 351
QY 527 EIPKWIENLAILKLSNNSFGNIPDELDCSLIWLIDNTNLNLTNGTTPAMFKOSK 586
DB 352 QIPQSLGRNLNLSVILSRNQSCHIPSISLANISLRLQDLISLNNFSGEIPV----- 403
QY 587 IANFIAGKRYVYIKNDGMKKECHGAGNLLEFGQIRSEQLNRLSTRNPNCTISRYVGH 646
DB 404 ----- 403
QY 647 SPFDNNGSMFLLDSYMLSGYIPKEIGSMPLYLTLNLGNDISGSDPEVDGLRLNT 706
DB 404 --SPDSQSLNLFNVSYNSLSGSP-----PLL----- 429
QY 707 LDISSNKLDGRIPQMSALTMLTEILDSNNNLGSPIPMGQFETPPAKFLNPGICGY- 765
DB 430 -----AKFNSSSFVGNILQCGYS 448
QY 766 PLPRC-----DPSNADGYAHHORSHGRRPASLAGSVAMGLLFSVCIFGLIIVG 814
DB 449 PSTPCLSQAPSGVIAPRPPEVSKHHHRKLSKTDILIIYAGVLLVLIILCCVLLFCLL- 507
QY 815 REMKRRKKEAELEMYABEGHNSGD-----RTANNWKLTGVXEA--LSINLAPE 865
DB 508 --RKRSTSK-----AGNGQATEGRAATWRTKGVPPVAGDVEAGGEGKLVHFD 556
QY 866 KPLRKLTPADLLQATNGFHNDSLIGSGFGDYKAILKOGSAVAIKLILHVSQGDREEM 925
DB 557 GPM-AFTADULLCAT-----AELMGKSTYGTIVYKALIEBGSQVAVRLEKIKGHRFE 610
QY 926 AEMETIGIKIRNLVPLGCKVG--DERLLVNEVWKYGSLEVDVQDRKGGVK--LKLS 981
DB 611 SEVSVLGKIRHPVNLALRAY-YLGPKEKLLVPDYMSKSLASFHGG--GGETFTIDWP 666
QY 982 TRRKIIAGSARGLAFLHNCSPHIIRDMKSSNVLLDENLEARVSDFGMARLMSAMDTL 1041
DB 667 TRMKIIADLARGLFCLH--SQENIIGHNLTSSNVLLDENTYAKIADFGSLRIMSTA-ANS 723
QY 1042 SVSTLAGTGVVPEVYOSFRCTGKYVSYGVVLELLTGKRPDSDPFGDNMLVGMVK 1101
DB 724 NVIATGALGYRAPBELSKLKANTKTIDISLGVILLETTRKSGVSNMGD--LPQWVA 781

```

Wed Mar 12 10:45:02 2003

us-09-823-394-2.rapb

Page 12

QY 1102 QHAKLR-1SDVFDELMKEDPALBIEILLQHKAVACLDPRMRPMTVMQAMFEKIQ 1159  
| : : | : : | : : | : : | : : | : : | : : | : :  
Db 782 SVVKEEMTNEVFADLMDASTVGDELLNTIKLALHCVDSPSPSKRPEVHQVLOQLEBIR 840

Search completed: March 10, 2003, 18:11:42  
Job time : 34.6693 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: March 10, 2003, 18:03:39 / Search time 25.6693 Seconds

(without alignments)  
1370.890 Million cell updates/sec

Title: US-09-823-394-2

Perfect score: 6183

Sequence: 1 MKTSSFFLSTVTLTFPSF.....GFSTIEMVMSIKVEEGLT 1196

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: Issued\_Patente\_AA:\*  
2: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6183	100.0	1196	4	US-08-881-706-2
2	1311	21.2	999	2	US-08-473-553A-5
3	1225.5	19.8	980	2	US-08-473-553A-6
4	1218.5	19.7	985	2	US-08-473-553A-2
5	1119.5	18.1	910	4	US-09-228-986-72
6	1087	17.6	1025	2	US-08-567-375-4
7	1087	17.6	1025	2	US-08-587-680A-4
8	1073.5	17.4	1012	2	US-08-475-881A-4
9	1057	17.1	947	4	US-09-228-986-73
10	891	14.4	1112	4	US-09-353-585-2
11	890	14.4	1016	4	US-09-180-439-8
12	889	14.4	1112	4	US-09-353-585-3
13	885	14.3	968	4	US-09-180-439-3
14	884	14.3	968	4	US-09-180-439-4
15	875.5	14.2	1023	2	US-08-475-881A-2
16	875.5	14.2	1023	2	US-08-567-375-2
17	875.5	14.2	1023	2	US-08-587-680A-2
18	875	14.2	1023	2	US-08-587-680A-2
19	798	12.9	630	4	US-09-228-986-76
20	796	12.9	630	4	US-09-228-986-71
21	795.5	12.9	630	4	US-09-228-986-71
22	795	12.3	638	4	US-09-180-439-6
23	715.5	11.6	805	4	US-08-228-986-74
24	705.5	11.4	863	4	US-08-945-983-2
25	673	10.9	666	4	US-08-666-271-2
26	668.5	10.8	655	4	US-09-228-986-68
27	668.5	10.8	659	4	US-09-228-986-70

28	655.5	10.6	707	4	US-09-228-986-80	Sequence 80, Appl
29	649.5	10.5	711	4	US-09-228-986-79	Sequence 79, Appl
30	629.5	10.2	523	2	US-08-473-553A-3	Sequence 3, Appl
31	589.5	9.5	903	4	US-09-228-986-78	Sequence 78, Appl
32	575	9.3	690	4	US-09-228-986-69	Sequence 69, Appl
33	524	8.5	858	2	US-08-265-628-2	Sequence 2, Appl
34	508.5	8.2	282	2	US-08-473-553A-4	Sequence 4, Appl
35	498.5	8.1	857	1	US-07-717-331F-3	Sequence 3, Appl
36	497	8.0	857	1	US-07-717-331F-2	Sequence 2, Appl
37	466	7.5	321	1	US-08-447-185-1	Sequence 1, Appl
38	445	7.2	739	4	US-09-503-922-1	Sequence 4, Appl
39	428	6.9	632	4	US-09-228-986-77	Sequence 77, Appl
40	384.5	6.2	501	2	US-08-980-060-6	Sequence 6, Appl
41	384.5	6.2	501	2	US-09-307-185-6	Sequence 6, Appl
42	377	6.1	712	1	US-08-587-689-2	Sequence 5, Appl
43	377	6.1	712	1	US-08-980-060-5	Sequence 5, Appl
44	377	6.1	712	4	US-09-307-185-5	Sequence 2, Appl
45	377	6.1	712	5	PCT-US96-09193-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-08-881-706-2  
Sequence 2, Application US/08881706  
Patent No. 6245965  
GENERAL INFORMATION:  
APPLICANT: Chong, Joane  
TITLE OF INVENTION: Receptor Kinase BIN1  
FILE REFERENCE: 07251/022001  
CURRENT APPLICATION NUMBER: US/08/881,706  
CURRENT FILING DATE: 1997-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 1196  
TYPE: PRT  
ORGANISM: Arabidopsis sp.  
US-08-881-706-2

Query Match	Best Local Similarity	100.0%; Score 6183; DB 4; Length 1196;
Matches 1196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY 1	MKTSSFFLSTVTLTFPSFSLFQASPSQGLYREIHQLISFKVLPDKNLLPDMSSNKN 60	
DB 1	MKTSSFFLSTVTLTFPSFSLFQASPSQGLYREIHQLISFKVLPDKNLLPDMSSNKN 60	
QY 61	PCTPGVTCRDKKTSIDLSKSKPLNVGSAVSSLLSLTGLESFLNSHINSVSGFKC 120	
DB 61	PCTPGVTCRDKKTSIDLSKSKPLNVGSAVSSLLSLTGLESFLNSHINSVSGFKC 120	
QY 121	SASLTSLDLSRNSISGPVTLTSLGSCGKFLNVSSNTLDPGKVSGLKNSLTVL 180	
DB 121	SASLTSLDLSRNSISGPVTLTSLGSCGKFLNVSSNTLDPGKVSGLKNSLTVL 180	
QY 181	SANSTSGANVGVWLSGCCGKFLKLAISGNKIGDDVDSRCVNLFLDVSSNFGTGF 240	
DB 181	SANSTSGANVGVWLSGCCGKFLKLAISGNKIGDDVDSRCVNLFLDVSSNFGTGF 240	
QY 241	LGDSCALQHDLDISGNKLSGDFSRALSTCTELKLNLSNQFVPIPLKSLQYISLAE 300	
DB 241	LGDSCALQHDLDISGNKLSGDFSRALSTCTELKLNLSNQFVPIPLKSLQYISLAE 300	
QY 301	NKFTGELPDLFSGACDTLTGLDLSGNHFGAVPPFGSCSLLESIALSSNNFSGELPMDT 360	
DB 301	NKFTGELPDLFSGACDTLTGLDLSGNHFGAVPPFGSCSLLESIALSSNNFSGELPMDT 360	
QY 361	LKKRGKLVLDLSEFNEFSGELPESLTNLSASLTLTDLSSNNSGPIPLKCNPKNTLOE 420	
DB 361	LKKRGKLVLDLSEFNEFSGELPESLTNLSASLTLTDLSSNNSGPIPLKCNPKNTLOE 420	

```

MEDMUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/473,553A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-60886/RFR/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 398-3249
TELEFAX: (415) 781-1969
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 999 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-473-553A-5

Query March 21.2%; Score 1311; DB 2; Length 999;
Best Local Similarity 32.4%; Pred. No. 3,96-87;
Matches 386; Conservative 161; Mismatches 392; Indels 254; Gaps 30.

QY 22 LSPQASPSQSLYREIHQHSFKQVLPD-KNLLPDMSSMKN--PCPPGVCV-RDDKVTST 77
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 11 LSTSLPLSLINDQATTLRQAKIGLSDPAQSSSSNDNDVTPCKNIGVSDATSNVSV 70
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 78 DLSSEKLVGSAVSSSLSLTGLSELPLSNHNGSVG--FKCASLTLSDSRSL 135
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 71 DLSSTMLVGPPL--LCHLPLSLSLSYNNSSINGSLADFDTCNLLSLDSENLV 127
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 136 GVVTLTSLGSCGKFLVNSNTLDPFGKVSGGLKLSLEVLDSANSLGAVVGMVL 195
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 128 GSIP-----KSLPEFLP----- 139
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 196 SPOGGEKHLMAISGNKISGVDVSRGVNLEFELDVSNNFSTGIP-FLGDSALQHLDISG 254
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 140 -----NKKFELISGNLSDTIPSSFGFRKLESLNLAG 172
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 255 NKLGDSPRAISCTELKLNLTSSNQF-VGPIPLP--PLKSLQVSLAKNKFTGEIPDFL 311
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 173 NPLSLTIPASLGNVTTLKEKLAVALNLPSPQISQGLTELVQVLMACNLVGPILPSL 232
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 312 SGACDTLTGLDSGNHFGAVPFPFGSCSLLESIALSSNNFSGELPMDTLKRGKLYLD 371
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 233 S-RLTSLVNLDTLFRQULGSTIPSWITQLKTYEQLELFPNNSFGELP--ESMGNNVTLKQFD 290
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 372 LSFNEFSGELPSLTLNLSLTLTDLDSNNFSGEILPNLQCNKNTLQELVYQNNGFTGK 431
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 291 ASNNKLTGKLPDLNMLL--NLESINLFPNNMLEGLPESTRS--KTLSEIKLEFNRLICV 346
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 432 IPPPLSSCSGLVSLHSFNYLSGTPSSLSGLSLKRLKLMNMLEGELPQELMYVTILE 491
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 347 LPSQLGANSPLQVYDLSYSTRFSGELIPAVVCEGKLEYLLILIDNSFSGEISNNLGRCKSLT 406
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 492 TLILDPLNDLGEIPSGLSNCTNLNMNLSLNNRNLGELPKMIGLENLALTKLNNFSFGN 551
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 407 RVRSLSNKKLSGQIPHGFWGIPRLSLLESLSDNSFGSLPKTLIGAKNLSLRISKRFSSG 466
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 552 IPDGLGCCRLLMLDLNTNLPNGTIPAMAFQSGKLIANFLAGKRYVYIRNDGKKECHG 611
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 467 IPNIGISLNGITELSGANDPSEGEIPESL----- 495
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 612 AGNLLFEGRSQNLRLSTRNCPNITSRYVGHTSPTEFDNNGSGMFLDASNNMLSGYIP 671
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 496 -----VTKQLSR-----LDSLKQVLSGEIP 516
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 672 KEIGSWPYFLINLGHNDISGIPDEVDLGLNLIDSSNKLDRIPQAMSAITMLTEI 731
DB 517 KRLRGKRNINLNNLANNHLSGIPREVGILPLVNLIDSSNQSGEIPLEQLNL-KLVNL 575
QY 733 DLNNNLGPIPEMGOFTFPKAPLNNPGLCGYPLPRCPDSNADGYAHHQSHGRPPAS 791
DB 576 NLSVNLHSGKIPPLVANNITY-AHPDIGNPGLC-----VDLDGCR-----K 615
QY 792 LAGSVAMG-----LIFSFVCIPLGLLVGEMRRRRRKEALEMYAAGHNSGDR 842
DB 616 ITRSNKIGYVILLITFILLAGIVFVGIVMFIACRKLRLAKSLSTLAA----- 663
QY 843 ANNTMKLTGVEALSTINLAEPKRLKTF-----ADLLQATNGFINDLSGGGFDV 897
DB 664 ---SKM-----RSFHLHPSHEHIDCLDEKN-----VIGFGSSGKV 697
QY 898 YKAILIKGSAVAIKKLIVHSGQGDRE-----FMAEMETIGIKHNLVPLIGYCK 947
DB 698 YKVELRGGEVAVAKKLNKSVKGGDEVSQSLNNDVPAALVEVLTGTRKSIYRLMCCS 757
QY 948 VGERLLVNEVMKGSLEDTYQPKKGVYKULSTRKKAISGARGIAFLHNCSPHIIH 1007
DB 758 SGDCCLVETWPGSLADYLAHGRKGVVIGWPERLRIALDAAGLSYLDHCVPIVH 817
QY 1008 RDMKSNVLDENLEARYSPFGMARL--MSAMDHLVSSTLACTPGVPEPYQSFRCST 1065
DB 818 RDVSSNILLDSQDYGAVDPGIAKVQMSGSKTPEMSGIGASCYIAPEYVYTLRVNE 877
QY 1066 KGDYISGVVILLITJRKRPDTPDFGDNILVGVV-KQHAKLRISSVPPELKEPDL 1124
DB 878 KSDIYSGVILLITVLTGKOPTDS-ELGDKDMAKVCTALDKGLEVIDPKL---DLKFK 933
QY 1125 IELLQHKVAVACLDDBAMRRPTMVOVMFKELIQAGSGIDSGSTRSIEDGG 1177
DB 934 EHSKVTIHIGLCTSPPLNRPKRVVIMLQVSGAVPCSSPNTSKSKTG 986

RESULT 3
US-08-473-553A-6
; Sequence 6, Application US/08473553A
; Patent No. 5859338
; GENERAL INFORMATION:
; APPLICANT: Meyerowitz, Elliot M.
; APPLICANT: Clark, Steven E.
; APPLICANT: Williams, Robert W.
; TITLE OF INVENTION: Plant Clavatal Nucletic Acids,
; TITLE OF INVENTION: Transformed Plants, and Proteins
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobach, Teet, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/473,553A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-60886/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 980 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-473-553A-6

Query Match 19.8%; Score 1225.5; DB 2; Length 980;
Best Local Similarity 29.4%; Pred. No. 6.5e-81;
Matches 352; Conservative 189; Mismatches 358; Indels 297; Gaps 35;

QY 14 LFFSPFSLSGFQSPSQSLREIHQSLSPDYV--PDKRLDPMSSNKP---CTPDGYT 68
DB 13 LHLVLFSPSPCP-----AYTMEVLLNKSWSVIGKGLHDMTHSSPDHACSPSGVS 65
QY 69 CRDD-KVTSIDLSKPLNVPFSAVSSLSLGLTESLFFSNHINGSVGFCSAISTSL 127
DB 66 CDDARVVISLNVSFPTPL---FGTISPEIGMLTHVNLTLAANNFTG-----EL 110
QY 128 DLSRNSLSGVTTLTSLGSCSGIKPLNVSSN--TLDPEKVSGLKLNLSLEVLDSANS 184
DB 111 PLENKSL-----TSIKVNLINNNGLNLGTPEGEI----- 139
QY 185 ISGANVGVWVLDGCGELKHLAISGNKISGDVDSRCVNLIELDVSSNNPSTGIFPLGDC 244
DB 140 -----LKAWDLEVLDTYNNFN----- 157
QY 245 SALQHLDISGNKLSGDSFRAISTCTELKLNLSNQFVGP1P--LPKSLQYSLAENK 302
DB 158 -----GKLPEMSELKTKLTSFGGNF 179
QY 303 FGEIPDPLSGACDILLGLDSGNHFGAVPFFGSCGLSLALS-SNNFSGELPMD-- 359
DB 180 FSGELPESTY-GIQSLEHYGLNGAGLSGKSPAPLSRLKNLEMYIGYVNSYTGVPREFG 238
QY 360 TLLKRGKLVLDLSPNPSGSELPESELTNLSASILLTLDSSNPFGLPLNLCQPKNTLQ 419
DB 239 GLTK--LEILMASCTLGLIELTSLNL-----KHLH 268
QY 420 ELYLQNNGTGKIPTLNSCELYSLHSFYVLSGTTIPSSLSGSLSKLDKLMMLMEGE 479
DB 269 TLFHLNNLTGHIPELGLVSLKSLDISINQGEIPQSPINLGNITLNLFPNNLYGQ 328
QY 480 IPOELMYVKTLETLIDPNDLGEIPELSGSLNCTLNMSLSNNRLTGEIKMYGLENTLA 539
DB 329 IPEALGELPKLEVEFVWEKNFTLQIPAVLGNNGNLIKLDVSDNHLTGLIPKDLCRGKLE 388
QY 540 IKLNNNSFGNIPDELDCSLIMLDINTNLFNGTIPAMFK-----QSGKIANFAGK 595
DB 389 MLILSNFFFGPIPELGLCKSLKIRIVKNLNGVPAAGLPLVTTIELDNPFSGE 448
QY 596 RYVYIKNDGMKKECHGAGNLEFQGISSEQLNRLSTNPNCTISRYVGGTSPFPNNGS 655
DB 449 LPVWISGVL-----DQYILSNMFSGEIPPAIGNFR 481
QY 656 IMFLLMSYMLSGYIPKEIGSPYFLINLGHNDISGIPDE----- 697
DB 482 LQTLUDLNNRFRGNIPREIFELKLSKINTSANNITGAGIDGDSISRCSTLISVLSNRIN 541
QY 698 -----VDLGLNLIDLSNKLDRIPQAMSAITMLTEIDLSNNNLGPIPEMGOFTFP 751
DB 542 GEIPKGINNVKVLGTNLISGNQLGSIPTGIGMTSLTTLTDSFNDLSGRVPLGGGFLV 601
QY 752 PPAKPLNPGLCGYPLR---CDPSNADGYAHHQSHGRPPASLAGSVAMGLIFSVCI 808
DB 602 NETSPAGNTYLC--LPHRYSC-PTRPGQTSDDHNTALSPSPSIVTV-----IAAIT 650
QY 809 GLIIVGERMRRRRRKKEALEMYAAGHNSGDRTANNVTKLVGEALSTINLAEPKRL 868
DB 651 GLILISVAIQMKKKKQK-----SLAMKLT-----AFQ--- 679
QY 869 RKLTF--ADLLQATNGFHNDSLISGGGFDVYKAILIKGSAVAIKKL-IHVSQGDREEM 925

```

Db 680 -KIDFSESEVLEC---LKENIIGKSGSIVRGSPNNVDAIKRLVORGGRSDHGT 735  
 Qy 926 AEMETIGIKHRNLYPLIGCKYGERLLVNEWMKXGSELDVLOPKKGVYKXSTRK 985  
 Db 736 AEIQTGRIRHRIYVLLGVANKDNTLLIYEMPGSIGELHSGKG--HLQWETHR 793  
 Qy 986 IAGSARGIAFLHNGSPHIIHRDMKSSNVLDENLEAVSDPGMARLMSAMDTLSVST 1045  
 Db 794 VAVEAKGICLYLHSDSPILHRDVSNMILLDSDEAHVADFGAKFLVDGAASECMS 853  
 Qy 1046 LAGTPGVPEYVYOSPCSTKGDVSYGVVLELLTGKRPDSDFGDN-NLVGWYKQHA 1104  
 Db 854 IASGYIAEYAVYTLKVDKSDVYSFGVLELLIAGKPVG--EFGEGVDIVRWVR-NT 910  
 Qy 1105 KAIRSDVPELMEKE--DPALE---IELLOHLKVAVACLDRAKRPMTVOYMA 1154  
 Db 911 EBEITQSDAIVAVAVDPRLTGVLTSVIHFKIAMCVBEBAARPTMRVYVM 966

LET 4

08-473-553A-2  
 ; Sequence 2, Application US/08473553A

Patent No. 5859338

GENERAL INFORMATION:

APPLICANT: Meyerowitz, Elliot M.

APPLICANT: Clark, Steven E.

APPLICANT: Williams, Robert W.

TITLE OF INVENTION: plant Clavatal Nucleic Acids,

TITLE OF INVENTION: Transformed Plants, and Proteins

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: Flehr, Hobach, Test, Albritton & Herbert

STREET: Four Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: United States

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08-473,553A

FILING DATE: 06-JUN-1995

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Silva, Robin M.

REGISTRATION NUMBER: 38,304

REFERENCE/DOCKET NUMBER: A-60886/RFT/RMS

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 985 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-473-553A-2

Query Match 19.7%; Score 1218.5; DB 2; Length 985;

Best Local Similarity 29.3%; Pred. No. 2.1e-80;

Matches 351; Conservative 189; Mismatches 359; Indels 297; Gaps 35;

Qy 14 LFFPSFSLSPQSPQSLYREIHOLISFMDVL--PDKNLLPWSMKNKP---CFPDGYT 68  
 Db 18 LHLVLFSPSCF-----ATTDEVLNLTSSMTGKGHLMHIISSSPDACHSPGVS 70  
 Qy 69 CRDD-KVTSIDLSKPLNVGSAVSSLSLTGLSIFLSNHSNGSVGFKCASLTSL 127

Db 71 CDDDAVIGLNVSTFPL---FGTISPEIGMLTHLVNLTLANNTG-----EL 115  
 Qy 128 DLSRNSISGEVTTLTLSGSCGKLPANVSN---TLDFPGKXSGGLKNSLEVLDSANS 184  
 Db 116 PLMEKSL-----TSLKVLNINSGNLTGTFFGEI-----144  
 Qy 185 ISGANVGVWVLSDCGELKHLAISGNKISGDVDSRCVNIIEFLDVSSNNFSGIPLGDC 244  
 Db 145 -----LKAMVLEVLDTYNNNNFN-----162  
 Qy 245 SALQHLDISGNKLSGFSRAISTCTELKLNLSNQPVGPIPE--LPKSLQYLSAENK 302  
 Db 163 -----GLPEMSELKKKLTSLFGSNF 184  
 Qy 303 FTGRIPLPLGACDTLTGLDLSGNHFGAVPPFGSCSLSEIALS--SNFSGELPMD-- 359  
 Db 185 FSGEIPESY-GDIQSLSYLUNGAGLSGKSPALSRUKLRMYITGYNSYTGVRREFG 243  
 Qy 360 TLKMRGLKVLDSFNEFSGELPESLTNLSASLTLDLSNNFSGIPLPMLCONPKNTLQ 419  
 Db 244 GLTK---LEILDMACTLTGTIPTSLSNL-----KHLH 273  
 Qy 420 ELYIQNNGFTGKIPTPTSSNCELVSLHSFNYSCTIPSSLSISKRLDKLMLMEGE 479  
 Db 274 TLFLHNNLTGHIPELISGLVSLKSLDLSINQLTGTIPQSFINGITLINLRNNLYGQ 333  
 Qy 480 IPOELMYKLTETLLDPNDLTGELPSGLSNCTINLVISNNRLTGEIPKMGILENTLA 539  
 Db 334 IPEALGELPKLIEVEWENNTTLDPLNLSGNGLIKLDVSDNHLTGILPDLORGEKLE 393  
 Qy 540 ILKLSNNSFSGNIDELGDCSLIWLNLNTNLNFGTIPAMFK-----QSGKLANFAGK 595  
 Db 394 MLIISNFFGPIPELQCKSKLTIRIVKNLNGVAPAGLPLVTIETLTPNFSGE 453  
 Qy 536 RYVYIKNDGMKKECHAGNLEFGQIRSEQLNRLSTRNPCITSRVYGHTSPFTDNGS 655  
 Db 454 LPVTSGDVL-----DQIYLSNNWSGELIPALGNPN 486  
 Qy 656 MMFLMSYNNLSGYIPKEIGSNPYFLINLGNHDSISIDE-----697  
 Db 487 LQTLPLDRRFRGNPREIPELKHLSRINTSANNITGIPDISISRCTSLISVDSLRNIN 546  
 Qy 698 -----VGLRGNTLIDLSNKLDRIPQMSALTMLEIDLSNNNSGPIPEMOQFET 751  
 Db 547 GEIPKINNKNVIGTINISGNQLTGSIPTGIGNMTSLTLLDSFNDLSGRVPLGQFLVF 606  
 Qy 752 PPAKFLNPNGLGYPLR---CDPSNADGYAHQSHGRAPASLAGVAMGLLFSVCIF 808  
 Db 607 NETSPAGNTYLC---LPHVSC-PTRPQOTSDHNTALFSPERIVTV-----IAAIT 655  
 Qy 809 GLILVGRNKRKRRKKEALEMYAEGHNSGDRITANNNTNKLTYKELASTILAPEKPL 868  
 Db 656 GLILISVAIRNNKKKNQK-----SLAMKLT-----AFQ---684  
 Qy 869 RKLTFF--ADLLQNTGPNHDSLIGSGFGDYKATIKDSAVAIIKL-I-HVSGQDREEM 925  
 Db 685 -KIDFSESEVLEC---LKENIIGKSGSIVRGSPNNVDAIKRLVORGGRSDHGT 740  
 Qy 926 AEMETIGIKHRNLYPLIGCKYGERLLVNEWMKXGSELDVLOPKKGVYKXSTRK 985  
 Db 741 AEIQTGRIRHRIYVLLGVANKDNTLLIYEMPGSIGELHSGKG--HLQWETHR 798  
 Qy 986 IAGSARGIAFLHNGSPHIIHRDMKSSNVLDENLEAVSDPGMARLMSAMDTLSVST 1045  
 Db 799 VAVEAKGICLYLHSDSPILHRDVSNMILLDSDEAHVADFGAKFLVDGAASECMS 858  
 Qy 1046 LAGTPGVPEYVYOSPCSTKGDVSYGVVLELLTGKRPDSDFGDN-NLVGWYKQHA 1104  
 Db 859 IADSYGIAEYAVYTLKVDKSDVYSFGVLELLIAGKPVG--EFGEGVDIVRWVR-NT 915  
 Qy 1105 KAIRSDVPELMEKE--DPALE---IELLOHLKVAVACLDRAKRPMTVOYMA 1154  
 Db 916 EBEITQSDAIVAVAVDPRLTGVLTSVIHFKIAMCVBEBAARPTMRVYVM 971

RESULT 5  
US-09-228-986-72  
Sequence 72, Application US/09228986  
Patent No. 6359198  
GENERAL INFORMATION:  
APPLICANT: Strabala, Timothy  
APPLICANT: Nieuwenhuizen, Niels  
TITLE OF INVENTION: Compositions isolated from Plant Cells  
TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signalling  
FILE REFERENCE: 11000/1020  
CURRENT APPLICATION NUMBER: US/09/228, 986  
CURRENT FILING DATE: 1999-01-12  
NUMBER OF SEQ ID NOS: 130  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 72  
LENGTH: 910  
TYPE: PRT  
ORGANISM: Pinus radiata  
-228-986-72

Query Match 18.1%; Score 1119.5; DB 4; Length 910;  
Best Local Similarity 30.5%; Pred. No. 3.1e-73;  
Matches 332; Conservative 166; Mismatches 365; Indels 227; Gaps 32;

QY 9 LSVTTLPFFSF---FSLSPQAS-----PSQSLREIHOLISTEDVL---PDKNLLPDMS 56  
DB 2 LLLATLSFILFLNPFAPSTPIAHPHPHPPRIITNADQHALAFKSAIYDPSQSLATSWL 61  
QY 57 SNNKPCFPGVTC--RDDKVTSIDLSKPLNPGFSAVSSLSLTGLESILFNSHINS 114  
DB 62 PNVSFQWGTGICRRRQVYISLNVSMGLQ---CTISPLANLSTLYVDLHN----- 112  
QY 115 VSGFKCA-----SLTSDSRNSLSGPVTTLTSLGSCGKFLNVSNTLDPFGKY 166  
DB 113 -NSFDCHIPYQGLTFLKMLRLSKQLOQSIIP--PTLANCSRLNLTLSFNNL--TGN 167  
QY 167 SGGI-KLNSLEVLDSLNSISGANVGVVSDCGSLKHLAISGKISDVVSCVNE 225  
DB 168 PPOCLLPNLCISLGINNTG-----TTPDCLNIS-----SLQ 202  
QY 226 FLVSSNNFTGIPF-LGDCSALOHLDISGNKLSGDFRAISTCTELKLNISNOQVGP 284  
DB 203 YLSLSQNLQSGVSEIGRLSQIIVDLFGNHLTGCIPESSLNCTNBLIDGNOQVGH 262  
QY 285 IP-----PLPKSLQVLSLANKTGTGIIPPFLGACDTLTGLDLSGNHFGAVPPFGSCS 340  
DB 263 IPSHLCTKKTQMLMYRLGANNQSSGVPSSLFN-CTKLOEIALPNQSLGIVPMELGKLT 321  
QY 341 LLESIALSSNNF-SG---ELPMDTLL-KMRGKAVLDLSNPSGSLPESLNTLSASLITL 395  
DB 322 HQRLPFGGNYFISGWTMCPILITLNSCSDIQYVDSLNNFTGQLPSPISGHSKRYHL 381  
QY 396 DLSNNFSPGPIPLNCONKNTLOELIYONNGFTKPIPTLSCSELYSLHSPVYLSGT 455  
DB 382 DLSNNELA-----GEIPPAIGNSSLTFNLGNNTFTGS 415  
QY 456 IPSSLGSLSKLDLKLMLNMLEGEIPOELMYKLTETLIDPNULTGTIPGSLNCTNIN 515  
DB 416 IPSSLIMQKLRILYDSDNNLNQGNIPMEIGQKSLIGLILYLSGNNLSGKIPDVANLQOLR 475  
QY 516 WLSLSNNRLTGEIPIKWIQLENLAILKLSNNSFSGNIPDELGDCRSL-IWLDLNTNLFNG 574  
DB 476 YLYLHNQSLGSDINANTLGCYVLLLDLSYNNLSCGHIOELAGLANLAFYNNLSNNLSG 535  
QY 575 TTPAMFKOSGKIAMPFLAGKRYVYIKNDGKKCHGAGNLLFQGIIRSQNLNLSRNP 634  
DB 536 HYPLELGR----- 543  
QY 635 CNIISRVTGHTSTPFDNNGSNMFLDWSYNNLSGYIPEKISGMYFLTLNMGHNDISGSI 634  
DB 544 -----FD---MLQAMIDISANGITGYIISIVSGMKEVAYIMLSTNALGPI 585

QY 695 PDEVDLGLNILLIDSSNKLDGRIPQMSALTMLTEIDLNNNLGPIPEMGQFETPPA 754  
DB 586 PVSISELSLQDLDLSSNNLSGCIPISLANLTMHLNFSFNLTSBGPVEGVFKXIGAT 645  
QY 755 KFLNPPGLCG--YPLPCDDPBNMGVYAHQRS--HGRPPASLAGSVAMGLFSEFVCIFGL 810  
DB 646 AFMGNLGLCGFWNLPPC-----YAHKRSVYLNLRKVIILVVVAIVL-----CLFLA 694  
QY 811 ILVGENMKR-RRKKEALEMYAEHGNSGDRYANNNTWLTGVKEALSNLAAPEKRL 869  
DB 695 ILMKRNCERNIQRDIGPSLNV--GH-----R 718  
QY 870 KLTFADLQATNGHNDLSLGGSGFDVYKAILKDSAVAIKKLHVSGQ-DREFMAEM 928  
DB 719 RISTAEVLATNEFSNDANLIGISFGKVKYGIINDGTWAV--KLNLQNEGAQKSPDRRC 777  
QY 929 ETIKIKRNLVPLGCKVQDERLLVNEVMKYSLEVDLQDPKKGVTXKSTRKIAI 988  
DB 778 KVLGRVHRNLIVITTCYSDLOIKALIFPLMPKSLDEKWLYPDDEGQSCINLQRIANTAI 837  
QY 989 GSARGLAFLHNCSPHIIHRDMKSSNVLLDENLEAVSDFGMARLMSAMDT--LSYST 1045  
DB 838 DIAQMTYLLHHGFQVYIHCDLKPNVLLGEMWTAVLIDFGIATICFANNEDGALTSTNA 897  
QY 1046 IAGTPGVPP 1055  
DB 898 LKSGTYIPP 907

RESULT 6  
US-08-567-375-4  
Sequence 4, Application US/08567375  
Patent No. 5952485  
GENERAL INFORMATION:  
APPLICANT: Ronald, Pamela C.  
APPLICANT: Wang, Guo-Liang  
APPLICANT: Song, Wen-Yuang  
APPLICANT: Szabo, Veronique  
TITLE OF INVENTION: Procedures and Materials for Confering  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/567,375  
FILING DATE: 04-DEC-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/004,645  
FILING DATE: 29-SEP-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/475,891  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/373,375  
FILING DATE: 17-JAN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Baecian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 023070-058930  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1025 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-567-375-4

Query Match 17.6%; Score 1087; DB 2; Length 1025;  
 Best Local Similarity 30.8%; Pred. No. 8.6e-71;  
 Matches 323; Conservative 140; Mismatches 371; Indels 216; Gaps 28;

QY 231 SNNFSGI--PFLGDSALQHDIDSGNKGSPRAISTCTELKLNINSSNOGVPIPL 288  
 DB 88 SSNLS-GIISPSLGNISFREDLGNVLSGELPELSLSKOLLESDNSIGSIFA- 145  
 QY 289 PLKSLQYLIAENKFTGEIPDFLSGACDTLTGLDSGNHFGAVPPFG-SCSLIESIAL 347  
 DB 146 -----AIGACTKLTSLDSHNOIRGMIPREIGASIKHLINLYL 183  
 QY 348 SSNNFSGELPMDTLLKRGKLVTDLSNERSGELPESLTNLSASILTLDSSNNFSGPIL 407  
 DB 184 YKNGLSGEIP-SALGNLTSIQEFDLSFNRLSGALPSSLQSL-SLTNNVLGQNNLSG-MI 240  
 QY 408 PNLQCN-----PKNTLQELYL-----QNNGFTGKIPTLSNCSBLV 443  
 DB 241 PNSTWNLSLRAFSVRNKGMLPTNAFKTILHLEVIMGNTNRFHKGKLPASVANAHLT 300  
 QY 444 SLHLSFNYLSGTPSPSLGSLSKLDLKLNM----- 475  
 DB 301 VIQYGNLFSGITTSFGRLENTLYAMRNLPQTRQDDWGFISDLTNCCKLQTLNLE 360  
 QY 476 --LSEGIQELMYKT-LETLILDNDLGEIPSGLSNCTNLMWISLNNRLTGEIPKWI 532  
 DB 361 NNLGVLNPSFSNLSSTLSFALBELKNTGSIPTDIGNLIGLQHYICNNFPGSLPSSL 420  
 QY 533 GLENTLILKLSNNSFSGNIPDELGDCKSLIWLDTNLNFGNTIPAAEFQSG---KIA 588  
 DB 421 GRKMLGILLAVENNLSSISPLAIGNLTENLILLGTNKSQWIPYLSMTLMLSLGSL 480  
 QY 589 ANFIAGKRYVYIKDGMKKECHGAGNLEFGIISQENLSTNPNCTISRYVGGHTSP 648  
 DB 481 TNNLSGP-----IPSELFNIGTSLIMNVSKNNLESGITP 515  
 QY 649 TDDNNGSMFLDMSYNTLSGYIPKEIGSPYLFILNIGNDISSIPDEVDLREGNILD 708  
 DB 516 EIGHLKNLVEFHAESNRISGKIPNTLGDQQLRYLYLQNNLLSGSIPALGQLETLTD 575  
 QY 709 LSSNKLGRIPQAMGALTMLTEIDSSNNNSGPIPEMGQFETPPAKFLNPGICG---- 764  
 DB 576 LSSNNLSGQIPTSLADITLHSLNLSFNSFVEGVPPTIAFAAGISISQAKKCGSIPD 635  
 QY 765 YLPREDPSNAGQYAHQHSHGRPASLAGSYAMGLFSPFCIFGLILVGRMKRRRKK 824  
 DB 636 LHLPRCCP-----LLENRKHFPVLPISVSLAALAILSSVILI-----TWKRTKK 682  
 QY 825 EAELEMYAEGHNGSGDRTANNNTWKLTVGKALSINLAAFEKPLKLTPLADLLQNTNGFH 884  
 DB 683 GADSRISMCH-----PL--VSYSQULVYATGIFA 709  
 QY 885 NDSLIGSGGFGDYKAILKDGSAVAIKLIIHSGQDREFFMAEMETIGKIKRNIVPLAG 944  
 DB 710 PTMLLSGSGFSGYKGLINIDHVAVKLVLENPKALSKFTBCALAKMRHRNLVKLV 769  
 QY 945 YCKVGDGR-----LVNEMVKYSLDVLQ---DPKKGVKLKLSTRKIALGSRGLAF 996  
 DB 770 ICSIDNRKNDFAKIYIDFNPNGSLDWMIHPTNDQADGRHLNRRVYITLIDVACALDY 829  
 QY 997 LHNHCSPHIIRDMKSSNVLLDENLEARYSDPGMARIM-----SADTHLSVSTLAGTGY 1052  
 DB 830 LHRHGPFPVACHDKSSNVLLSDDMVAHVGDFGLARIILVDGTSLLIQOSTSSNGFITGTY 889

QY 1053 VPEYIOSFRCTGDIYVYGVVLLLELTGKRPDS---PDFGDNMLVGMVQKHAHL--- 1106  
 DB 890 AAPGVGLASTHGDYISGLVLEIVTSKRPDS-----LQYVELGLH 942  
 QY 1107 -RISVDPPELM-----KEDPALIE--LLOHLKVAACLDL-RAMRPYVQYM 1152  
 DB 943 GRVDVVDTKLIDSENMWNTNNSPCRRITECTIVMLRLGLSCQELPSSRPT----- 997  
 QY 1153 AMFEKIQAGSGIDSGSTIRSDGGEFSTIE 1182  
 DB 998 -----GDIIIDELIMIKONLSGLFVCE 1019

RESULT 7  
 US-08-587-680A-4  
 Sequence 4, Application US/08587680A  
 Patent No. 5977434  
 GENERAL INFORMATION:  
 APPLICANT: Ronald, Pamela C.  
 APPLICANT: Wang, Guo-Liang  
 APPLICANT: Song, Wen-Yuang  
 APPLICANT: Szabo, Veronique  
 TITLE OF INVENTION: Procedures and Materials for Confering  
 TITLE OF INVENTION: Disease Resistance in Plants  
 NUMBER OF SEQUENCES: 27  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Townsend and Townsend and Crew LLP  
 STREET: Two Embarcadero Center, Eighth Floor  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94111-3834  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/587,680A  
 FILING DATE: 17-JUN-1996  
 CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/373,375  
 FILING DATE: 17-JAN-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/475,891  
 FILING DATE: 07-JUN-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/004,645  
 FILING DATE: 29-SEP-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/567,375  
 FILING DATE: 04-DEC-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Baebian, Kevin L.  
 REGISTRATION NUMBER: 34,774  
 REFERENCE/DOCKET NUMBER: 023070-058940US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 576-0200  
 TELEFAX: (415) 576-0300  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1025 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-587-680A-4  
 Query Match 17.6%; Score 1087; DB 2; Length 1025;  
 Best Local Similarity 30.8%; Pred. No. 8.6e-71;  
 Matches 323; Conservative 140; Mismatches 371; Indels 216; Gaps 28;  
 QY 231 SNNFSGI--PFLGDSALQHDIDSGNKGSPRAISTCTELKLNINSSNOGVPIPL 288



```

Db      88  SSNLS-GTISPSLGNLFLRLDGNVLSGELPPELSRLQLLELSDNSIQGSIIPA- 145
Qy      289 PLKSLQVLSLAENKFTGEIPDPLSGACDTLTGLDLSGNHFYAVPPFG-SCSLESIAL 347
Db      146  -----AIGACTKLTSLDLSHNLQARGIPREIGASIKHLSNLT 183
Qy      348 SSNNFSGELPMDTLTKMRGKLVLDLSFNEFSGELPSLNTLSASLITDLSNNFSGPIL 407
Db      184 YKGLSGEIP-SALGNLTSIQEFDLSFNRLSGAIPSSLGQLS-SLITNMLGQNNLSG-MI 240
Qy      408 PNLCON-----PKNTLQELYL-----QNNGFYTGKIPPLSNCSLEY 443
Db      241 PNSTWNLSSLRARSVRNKLGMIPNNAFKTLHLLEVTDMGNTRFGKIPASVANASHLT 300
Qy      444 SLHSFVYLSGTIPSSLSLSKLRDLKMLNM----- 475
Db      301 VIQYGNLFGIITSGFGRRLNLTLEYLWRNLQTRBQDQWPFISDLTNCSKLTQTLNIGE 360
Qy      476 --LEGELPQELMYKT--LETLIDPNDLTGELIPSGLSNCTNMLNLSNNRLTGEIPKI 532
Db      361 NNTGVIPLNSFSLSTLSFLALEINKTIGSIPKQIGNLIGLQHLVLCNNFRGSLPSSL 420
Qy      533 GRLENTALIKLNNFSGNIPDELGDGRSLIMLDNTNLFNCTIPAAFKQSG-----KIA 588
Db      421 GRLLKNGILAYENNLSGSLPLAIGLTELINLLGCTNKFSGMITYTTLNLTSLGLS 480
Qy      589 ANFIAGKRYVYINDQMKKECHGAGNLEFGQIRSEQNLRLSTRNPNCTITSRYVGHSTP 648
Db      481 TNNLSGP-----IPSELFNIGLTSIMINVSNNLEGSIPQ 515
Qy      649 TFDNNSMFMFLDMSYMLSGYIPKEIGSMPLYFILNGLNDSSGIPPEVDLGLNLTLD 708
Db      516 EIGHLKNVLFPAHSESRSLGKIPNTLGDCCQLRYLYLQNNLSGSLPQALQQLGLTLD 575
Qy      709 LSSNKLGRIPQAMSAITLWTEIDLSNNNLSGPIPEMGOFETFPAPKPLANNPGCG----- 764
Db      576 LSSNNLSGQIPSTLADITWLSHLNLSFNSFVEVPTTGAFAASGISIQNAKLCGGIPD 635
Qy      765 YPLPRCDPSNAGVYAHQHSRPPASLAGSVAMGLFSPVCIFGLILVGREMKRRRK 824
Db      636 LHPRCPCP-----LLENRGHPVLPISVSLAALAILSLYLLI-----TWKKTKK 682
Qy      825 EAELEWYAEHGNSGDRYANNNTNMLTGVKALSLINLAPEKPLKLTFFADLLQATNGFH 884
Db      683 GAPSRTRMKH-----PL--VSYSQVYATDGF 709
Qy      885 NDSLIGSGFGDVYKAILKDGSAVAIKLHVSGQGRFPAEMETIGIKHRLVPLLG 944
Db      710 PYNLLSGSGFSGVYKGLNIPDHVAVKYLKLENPKALKSFAECEALRNRRHRLVKIYV 769
Qy      945 YKVGDER-----LLVNEVMKYGSIEDVLO--DPKKGCVKLTSTRKIALISGARLAF 996
Db      770 ICSSIDNKGNDPKALVYFPMFNGSLIEDWIPETNDQADQRLNHRHRTVILLDVAACLDY 829
Qy      997 LHHNCSPIIHRDMKSNVLLDENLEAVSDFGMARLL-----SAMDTHLSVSTLAGPGV 1052
Db      830 LHRHGPPEPVVHCDIKSNVLLDSDMVAVHVGDFGLAKILVDGTSILQOSTSMSGPITGTG 889
Qy      1053 VPPEYQSGFSGTAKDVYSYGVVLELLTGRPTDS--PDGDDNNLVGVWQKHAU- 1106
Db      890 AAFEGYGLIASHTHGDIVSYGILVLEIYTGKRPDTSFPRDLG-----LRQYVELGHA 942
Qy      1107 -RISDVPEELM-----KEDPALETE--LLQHLKAAVACLD--RAMRFPYVQW 1152
Db      943 GRVYDVVDTKILIDSEKWLNTNNSPCRRITECTIVMLRLGLSCOSLPSSRPT----- 997
Qy      1153 AMFKEIQAGSGIDQSOTRSIEDGFSSTIE 1182
Db      998 -----GDIIDELINAIKQNLGLFPVCE 1019

```

```

US-08-475-891A-4
; Sequence 4, Application US/08475891A
; Patent No. 5859339
; GENERAL INFORMATION:
; APPLICANT: Ronald, Pamela C.
; APPLICANT: Wang, Guo-Liang
; APPLICANT: Song, Wen-Yang
; TITLE OF INVENTION: Procedures and Materials for Confering
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,891A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; FILING DATE: US 08/373,375
; ATTORNEY/AGENT INFORMATION:
; NAME: Baselian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 02370-058910US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1012 amino acids
; TYPE: amino acid
; STRANDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..1012
; OTHER INFORMATION: /note="Xa21 Xanthomonas spp.
; OTHER INFORMATION: disease resistance polypeptide RRK-B
; OTHER INFORMATION: from rice (Oryza sativa)"
; US-08-475-891A-4

Query Match 17.4%; Score 1073.5; DB 2; Length 1012;
Best Local Similarity 30.5%; Pred. No. 8.2e-70;
Matches 320; Conservative 139; Mismatches 362; Indels 229; Gaps 29;

```

```

Qy      231 SSNNFSGI--PFLDGCALQHLDISGNKLSGDFSRASITCTELKTLNLSNQFVGIPPL 288
Db      88  SSNLS-GTISPSLGNLFLRLDGNVLSGELPPELSRLQLLELSDNSIQGSIIPA- 145
Qy      289 PLKSLQVLSLAENKFTGEIPDPLSGACDTLTGLDLSGNHFYAVPPFG-SCSLESIAL 347
Db      146  -----AIGACTKLTSLDLSHNLQARGIPREIGASIKHLSNLT 183
Qy      348 SSNNFSGELPMDTLTKMRGKLVLDLSFNEFSGELPSLNTLSASLITDLSNNFSGPIL 407
Db      184 YKGLSGEIP-SALGNLTSIQEFDLSFNRLSGAIPSSLGQLS-SLITNMLGQNNLSG-MI 240
Qy      408 PNLCON-----PKNTLQELYL-----QNNGFYTGKIPPLSNCSLEY 443
Db      241 PNSTWNLSSLRARSVRNKLGMIPNNAFKTLHLLEVTDMGNTRFGKIPASVANASHLT 300
Qy      444 SLHSFVYLSGTIPSSLSLSKLRDLKMLNM----- 475

```

Db 301 VIOYGNLFGSGLITSGRGLRNLTELYMNLFPOTREODMGFISDLTNCCKLQTLNIGE 360  
 Qy 476 --LEGEPOSLMVKT--LETLILDNDLGEIPSGLSNCTNLMWISLNNRNLGEIPKMI 532  
 Db 361 NNNGGVLPNSFSLSTSLFLALELNKTSIKDIGNLIGLOHLVLCNNRFGSPSL 420  
 Qy 533 GRLENTLALIKLNNNSFSGNIPDEIGDCRSILWLDLNTLNGTIPAMFKOSG---KIA 588  
 Db 421 GRKLKGLILAYENNLSSGIPLAIGNLTELNTLLGTNKSFWIPYTLSTLTLSTLGLS 480  
 Qy 589 ANFIAGRRVYIKNDGKKCKGAGNLEFGIRSEDLNLSIRNPNCTISRYGGHTSP 648  
 Db 481 TNNLSGP-----IPSELENTQTLSTIMTWSKNLBSITPQ 515  
 Qy 649 TFDNNGSMFLDMSYNNLSGYIPKEISMPYLFILNGLNDISGIPDEVDLRLNILD 708  
 Db 516 EIGHKLVLFHNASNLSSKIPNTLGDCCOLRLVYQNNLLSGSIPALGQKGLETLD 575  
 Qy 709 LSNKLDGRIPQMSALTMLTEIDLNNNLSGPIEMGOFETPPAKFLNPGLCG---- 764  
 Db 576 LSSNNLSGQIPSTSLADITMLSLNLSFNSFVGEVPTIGAPAAAGSISIQNAKLCGIPD 635  
 Qy 765 YPLRCPDSNADGVAHHQSHGRRPASLAGSVAMGLFSPVCIPGLILVGRERKKRKK 824  
 Db 636 LHLPRCCP-----LLENKHPVLPISVSLAALAILSLYLTL-----TMMKRTTK 682  
 Qy 825 EAELMYAEHGNSSGDRYANNTNMLTGVEKALSINLAPEKPLRLKTLFADLIQATNGFH 884  
 Db 683 GAPSTSKGK-----PL--VYSQVLVATDGFA 709  
 Qy 885 NDSLIGSGFSDVYKAILKDSAVAIKKLIVSGGDRFEMAEWETTKIKHRLVPLLG 944  
 Db 710 PTNLIGSGFSFSGVYKGLKLNPKAL-----KSFPAECBALNNHRLVPLVIT 756  
 Qy 945 YCKVDEER-----LLVNEVMYKGSLEVDLQ---DEPKGGVKLKLSRKRKIAIGSARGIAF 996  
 Db 757 ICSSIDNKGNDPKAIVYDFMPNGSLJEDWIHPETNDQADQRLNHRRTYLLDDVACALDY 816  
 Qy 997 LHHNCSPHIHRDMKSNVLLDENLEARVSDFGMARLM---SAMDTLSTVSLTAGTPGY 1052  
 Db 817 LHRHPEPVPVHCDIKSSNVLLDSMDVAHGDGFLARILVDTGSLIQOSTSMNGFITGTY 876  
 Qy 1053 VPEYYSRGTCTKGDVYSXGVVLELLTGKRPDS---PDEGDNVLGVWYKONAKL--- 1106  
 Db 877 AAPEYGVGLIAGTHGDVYSGLVLEIVGKRPDSTRPDLG-----LRQYVELGLH 929  
 Qy 1107 -RISDVFDEPLM-----KEDPALEIE--LLQHLKVAVACLD--RAMERPPTVQVM 1152  
 Db 930 GRVTDVVDTKLILDSNNMLNSTNNSPCRITBECIYWLTLGLSCQELPSSRTP----- 984  
 Qy 1153 AMFEIQAGSGIDSGTIRSIEDGFSSTIE 1182  
 Db 985 -----GDIIIDELNAIKONLSGLFPVCE 1006  
 RESULT 9  
 US-09-228-986-73  
 ; Sequence 73, Application US/09228986  
 ; Patent No. 6359198  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Strabala, Timothy  
 ; APPLICANT: Nieuwenhuizen, Niels  
 ; TITLE OF INVENTION: Compositions isolated from Plant Cells  
 ; TITLE OF INVENTION: and their Use in the Modification of Plant Cell Signalling  
 ; FILE REFERENCE: 11000/1020  
 ; CURRENT APPLICATION NUMBER: US/09/228,986  
 ; CURRENT FILING DATE: 1999-01-12  
 ; NUMBER OF SEQ. ID NOS: 130  
 ; SOFTWARE: PASCSEQ for Windows Version 3.0  
 ; SEQ. ID NO 73  
 ; LENGTH: 947  
 ; TYPE: PRF  
 ; ORGANISM: Pinus radiata

US-09-228-986-73  
 Query Match 17.1%; Score 1057; DB 4; Length 947;  
 Best Local Similarity 30.5%; Pred. No. 1,2e-68;  
 Matches 331; Conservative 164; Mismatches 377; Indels 214; Gaps 36;  
 Qy 137 PVTLLTSLGSC--SGLELVNVSNTLDPFGKVGSLKNSLEVLIDLSANSIGANVGVY 194  
 Db 9 PFLSLPLTSLDSBGTCLDQOQYVMK-----KLSYTP---SWTVKSDNPGGWS 56  
 Qy 195 LSDGCGE---LKHLAISGNKISGPDVDSRCVNLFLDVSSNNSTGIPFGDSCALQHL 251  
 Db 57 GVNCTAERSNVTLEHMSGFRMKGNAQOTIC-----KLQALQVLD 95  
 Qy 252 ISGNLSGDPFRAISTCTELKLNLSNNQFVG--PIPL-PLKSHQVLSAENKFTGEIPD 309  
 Db 96 VSDNLSTPSDNDIOACTNLPLSLNLSNPLPGSSLPBLAPRKHLFDVHNGPAGFEP 155  
 Qy 310 FLGACDTLTGLDLSGNHFGVAVPPFGSGSLBSLASSNNFSGELPMOTLKMRLKV 369  
 Db 156 QIQHLLD--LRVLNLTNNFSGPIPSFLGNLTTLTEKIDFSGQYFGEPEPK--LVACTSLTY 213  
 Qy 370 LDLSFNEFSGELPESLNTLSLTLTDLSSNNFSGPI-----LPNIQNPXKTLQELYL 423  
 Db 214 LDLSFNRRLTGOIPDNI SNL--IHLFTLLSSNNLTGTLPKTLDRLVNLTFRASN----- 265  
 Qy 424 QNNGFTGKIPPTLSNCSLEVSILHSFNYSGLTIPSSLSLSKLDLKLMLMGEIPQ-- 482  
 Db 266 -KQQLIGRIPVOLAKLTLEHFLDLSYNGLNFTIPELPALSNLQTLDTLTKNLIGEIPQN 324  
 Qy 483 ---ELMYVKTLETLI-----LDEND--LGEIPSGLSNCTNLMWISLSN 521  
 Db 325 FSRKILRLRIGQNLKGNIPLTIGMSMLTYLEMNDNSLDQIPLQVNCIKLOLDLGN 384  
 Qy 522 NKLTEGIPKTRGLNMLAIKLSNNSFSGNIPDEIGDCRSILWLDLNTLNGTIPAMF 581  
 Db 385 NNLSSGLTNQLPSLQLOVYKLKNNNFVGSIPYILSSSNLSYDLSNTLNGSIPSNIF 444  
 Qy 582 KQSGKIAANFIAGKRYVYIKNDGKKCKGAGNLEFGIRSEDLNLSIRNPNCTISRV 641  
 Db 445 -----NLSKQNLRL--QNNKLTGAILPNTV----- 467  
 Qy 642 YGHTSPPTDNNGSMFLDMST--NMLSGYIPKEISMPYLF--LNLGNDISGIPDEV 698  
 Db 468 -----GSGVYLELQLGNNLTGTMPLTEIGPARKLOIQNLSCNSLGGIPLNTL 516  
 Qy 699 GDLRGNIILDSNKLKDGRIPOMSALTMLTEIDLNNNLSGPIEMGOFETPPAKF-- 756  
 Db 517 SGLYMLIIDLNNKLTGVPGLTLMLSLTILNLSNLSLTGLPK-----FPMSTSL 570  
 Qy 757 ---LNNPGLCGYPLRCPDSNADGVAHHQSHGRRPASLA-----GSVAMGLLF 802  
 Db 571 IILDGTNPGTLA-----GQGSAPASASARKKISALLIIGVAVAGAVF 612  
 Qy 803 SFVCIIFLLIVGEM--RKRKKKEALEMYAEHGNSSGDRYANNTNMLTGVEKALSIN 860  
 Db 613 AIVAV--GLFTVASKYFGRDQDQPEVOLARKIEGHTHPDSI-----HRLRD 659  
 Qy 861 LAPEKPLRLKTLFADLIQATNGFHNDSLIGSGGFGVDVYKAILKDSGSAVAIKKL-----IH 915  
 Db 660 ---FEKVEA-----TLDPANVFLKXK-----FSTYYKAVMPSGISYGVKLNMSDRIF 705  
 Qy 916 VSGGDRFEMAEWETTKIKHRLVPLLGCVKYGDERLLVNEPMKGSLEVDLQDPKKG 975  
 Db 706 KSG--SYRKLGALEKQKGLKHPNLPPLAHVLDTSAYLFEYEVHNGSISEFHTSNVS- 763  
 Qy 976 VKLSTSRKIAIGSARGIAFLHNNCSPHIHRDMKSNVLLDENLEARVSDFGMARLMS 1035  
 Db 764 -VLDMPSCRRLAIGVAGGLAFL--HGCHP1PHLDITTKNLLSLNRPQGLDEICKVD 821  
 Qy 1036 AMDTHLSTVLAGTPGVVPEYYSFRCTKGDVYSXGVVLELLTGKRPDSDPFGDN 1095  
 Db 822 PSKSTGISALAGSVGVVPEYVATYMTAVTAAANVSFGVILTEHLGRTPTIS-----GMD 877

QY 1096 LVGWYKQ--HAKLRISDVDPDELMKEDPALEIETELIOLIKVAVACIDDPAMRPTVVOYMA 1153  
DB 878 LAKWOSTLSGEEWEQILDTGIRNFVSVOIQWEMIAMLKVALSCVSSPESRPRKRVVVG 937  
QY 1154 MEKEIQ 1159  
DB 938 MLOWVR 943

## RESULT 10

US-09-353-585-2  
Sequence 2, Application US/09353585  
Patent No. 6287865

## GENERAL INFORMATION:

APPLICANT: Dixon, Mark S  
Jones, David A

TITLE OF INVENTION: Plant pathogen resistance genes and uses  
thereof

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: Nixon & Vanderhye PC  
STEER: 8th Floor, 1100 No. 6287865th Glebe Road  
City: Arlington  
State: Virginia  
Country: United States of America  
ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/353,585  
FILING DATE: 15-Jul-1999  
CLASSIFICATION: C12N 15/29, 15/82, A01H 5/00, A01N 65/00, C12Q  
1/68

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/930,277  
FILING DATE: 27-OCT-1997  
APPLICATION NUMBER: PCT/GB96/00785  
FILING DATE: 01-APR-1996  
APPLICATION NUMBER: GB 9506658.5  
FILING DATE: 31-MAR-1995

ATTORNEY/AGENT INFORMATION:  
NAME: Ms Mary J Wilson  
REGISTRATION NUMBER: 32,955  
REFERENCE/DOCKET NUMBER: 620-69  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816-4000  
TELEFAX: (703) 816-4100  
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1112 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Tomato

SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-353-585-2

Query Match 14.4%; Score 891, DB 4; Length 1112;  
Beet Local Similarity 30.6%; Pred. No. 1.9e-56;  
Matches 283; Conservative 129; Mismatches 329; Indels 184; Gaps 30;

QY 74 VTSIILSPFLNWSAVSSLSLTGLESPLNSHNGSVSGPK-SASITSIDLSRN 132  
DB 217 LTELDSIDNAN---GSIFASLGNMNNISFLPLNGNQLSGSIPETICYLRSILYIDLSN 273

QY 133 SLSPVYTLTSLGSGSGELKELNVSNTLDFPGKVSGLK-----LNSLEVDLSANSISG 187  
DB 274 ALNGSIP--ASLGNLNNLSFLFLYGNQL-----SGSIPETICYLRSILYIDLSN 325  
QY 188 ANVVGWVLSDGCGELKHLA---ISGNKISGDVDS--RCVNLSEFLDVSNNSTGIP-FL 241  
DB 326 S-----IPASIGNLKNLSRLNLVNNQSLSGSIPASLGNLNNLSMLYLYNNQSLSGSIPASL 379  
QY 242 GDCSALQHDIDSGKLSGDPSPRAISTCTELKLNLSNQFVGPPIPL--PLKSLQYLSLA 299  
DB 380 GNINLNNLSMLYLYNNQSLSGSIPASLGNLNNLSRLYLYNNQSLSGSIPETICYLSTLYL 439  
QY 300 ENKFTGEIPD-----FL-----SGACDTLTGDLSGNHYGAVPEPF 336  
DB 440 NNSINGFIPASFGMNSNAFLFLYENQSLASVPEITIGLRSILNVLDISENLANGSIPASF 499  
QY 337 GSCSLIESIALSSNNPFGELPMDTLKNGKLVLDLSFNEFSGELPESLTMLSASLTLD 396  
DB 500 GNINLNNLSRLNLVNNQSLSGSIP--EIGYLRSILNVLDISENLANGSIPAFGNLN--NLSRLN 557  
QY 397 LSSNPSGPIPLNLCNPKNTLQELYLQNGFTGKIPTLSNCELYSLHSFVYLSGTI 456  
DB 558 LVNNQSLSGSIPETI--GYLRSILNDLSENLANGSIPASLGNLNNLSMLYLYNNQSLSGSI 615  
QY 457 PSSLSLSKLRDLKLMNLMEGELPQELMYVKTLETLIDFNDLTGELPQGLSNCTNLNW 516  
DB 616 PEIIGYLSLTYLTLSSGNLSNGLIPASFRANKNLQALILNANNLIGELPSSVCNLSTLEV 675  
QY 517 ISLSNNRLTGEIPKWTIGLENLAILKLSNNSFGSNIPELDCBSLTMLDLNTMIFNGTI 576  
DB 676 LYMPNNLKGVYPOCLGNISMLQYLSMSNSNPSGELPSSISNLTSLQILDGRNNLEGA 735  
QY 577 PAAM-----FKOSGKIAANPIAGKRYVYIKNDGKKE-----CH----- 610  
DB 736 PQCFNNSISLEVFMNKNKSLGTLPTNPSIGCSLISLNGNELEDEIPSLDNCCKLQV 795  
QY 611 -----GAGNLTEFGQIR--SEQL-----NRLST 631  
DB 796 LDLGNQNDLTFPMWGLTPELRLVRLTNKTLHNPIRSSRAELMFPDLRTIIDSRNAPSQ 855  
QY 632 RNPCKIISRYVGGHT-----SPTFDN--NGSMW-----FLDSYNN 665  
DB 856 DLPTSLFEHLKGMRTVXKMBEBSYVDSVYVYVYKGELEIVRLISLYTYIDLSNK 915  
QY 666 LSGYIPKEIGSMPLFTILNGLNDISGIPDEYDGLRGNIILDSKDKLGRIPQMSAL 725  
DB 916 FECHTPEVLDLIRILNLVSHNLQYIPSSLSLSLSDLSFQSLSGELPQOLASL 975  
QY 726 TMLTEIDLSNNNLSPLENGQETFPFAKFLNPPGLCGYPLPR--C--DPSNADGYA--- 779  
DB 976 TFLFELNLSHNYLQGCIPQGFRTFESNGBEGNDGLRGYFVSKGCKXDVSKRYTVA 1035  
QY 780 -HHQRSHGR-----RPASLAGSVAMGLIFSVICIFGILVY-----REMRK 820  
DB 1036 LBDDESNSEFENDFWKXALWG--YSGGICIGISMTYILISGNLRLARLIEKLEHKLIMQ 1094  
QY 821 RKKKAELEMYAGHGNSGDRYANN 845  
DB 1095 RKKQ-----RQGRNRYRRNN 1110

## RESULT 11

US-09-180-439-8

Sequence 8, Application US/09180439  
Patent No. 6225532

## GENERAL INFORMATION:

APPLICANT: Dixon, Mark S  
Jones, David A

TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof  
FILE REFERENCE: 620 - 53

CURRENT APPLICATION NUMBER: US/09/180,439  
 EARLIER FILING DATE: 1998-12-06  
 EARLIER APPLICATION NUMBER: PCT/GB97/01249  
 EARLIER FILING DATE: 1997-05-08  
 EARLIER APPLICATION NUMBER: GB 9609661.3  
 EARLIER FILING DATE: 1996-05-09  
 EARLIER APPLICATION NUMBER: GB 9619924.5  
 EARLIER FILING DATE: 1996-09-24  
 NUMBER OF SEQ ID NOS: 10  
 SOFTWARE: Patent In Ver. 2.0  
 SEQ ID NO: 8  
 LENGTH: 1016  
 TYPE: PRF  
 ORGANISM: Lycopersicon esculentum  
 US-09-180-439-8

Query Match 14.4%; Score 890; DB 4; Length 1016;  
 Best Local Similarity 29.0%; Pred. No. 2e-56;

Matches 304; Conservative 146; Mismatches 352; Indels 248; Gaps 38;

2 KTESFPLVTTLPFFSPFSLSPQSPQSLYREIHQIISFKDVLDPKN--LLPWSNK 59  
 7 KVFSS--LQFTVY--LFTVAFST-----EATALLKMKATFKQNNSLASWTSS 56  
 60 NPC-TFDGVTGRDRTYSLDSSKPLNVGSAVSSLL-----SLGLESLFSLNSHNG 113  
 57 NACDWCWVCLNGRVNT-----LNTNASVIGTLVAFPFSSLPFLNLDLNNISG 109  
 114 SV-----SGF-----KCSASL 124  
 110 TIPEIGN/MLVYLDLNTNOISGTTIPQISLAKLQIIRFNHNLGFTPEIGYRSL 169  
 125 TSLDSSRLSGPVTTLTSLGSCGKLFNYSNTLD-----F 162  
 170 TKLSIGINFLSGIP--ASLGNMTLSFLFYENQLSGFTPEIGYRSLTKLSLDINFL 227  
 163 PGKVSGL-KLNSLEVLDSANISGA--NVGVVLSGCCGELKHLAISGKISGDVDS 219  
 228 SGSI PASLGNLNLISFLIYNNQLSGSIPEIGYRSL-----LTKLSIGINFLSGIPAS 282  
 220 --RCNLEFDVSSNNFSGTIP-FLGDCSALOHDISGKLSGDFSRASITCTELKINI 276  
 283 LGNINLRLDLYNNKLSGSIPEIGYRSLTYLDLGENALNGSI PASLGNLNLPMYXL 342  
 277 SSNOFGVPIPU--PLKSLQVLSLAENKFTGEIPDLGACDTLTGLDLSGHHFYGAVP 334  
 343 YNNQSGSIPEIGYRSLTYLDLGENALNGSI PASL--GNLNLRLDLYNNKLSGSIPE 401  
 335 FGSGCSLESLASSNNGEELPMDTLKMRGLKVLDSFNEFSGELPESLTLNLSASLT 394  
 402 EIGYRSLTYLDLGENALNGSI PASLGNLNLPMYLYNNQLSGSIPEIGYRSL--SLTE 459  
 395 LDISNNGSGPI--LPNLQNPXTLQELYLNNGFTGKIPPTLNCSELSVLSHLSFNY 451  
 460 LHYGNNSLNGSI PASLGNL--NMLPMYLYNNQLSGSIPEIGYRSLTEFLGANS 514  
 452 LSGTIPSSIGSLKLDLKLMLMTEGELPQBLMYKTLTETLLDPNDLTGEIPBSGLNC 511  
 515 LNSGIPASLGNLNLRLYLYNNQLSGSI PASFGNNRNLTQLFLSDNDLIGIPFVCL 574  
 512 TNLNWLSTLNNRLTGEIPKMGIGLEMLALIKLSNNSFGSNIPDELGDCLSLTMDLNTYL 571  
 575 TSLLEVLYMKNKLGKVPQCLGNISDHLISMSNSFRGELPESISNLTSLKILDPGRNN 634  
 572 FNGTIP-----AAMF-----KSGKLAANEIAGRYVYI-----KNDGM 605  
 635 LBGALPQFPFNISLQVFMQNNKLSGTLPTNSIGSLSLNHLGNELADELIPSLDNC 694  
 606 KK-----ECHG-----AGULLPFGGRSSQANR 628  
 695 KKLQVLDLGNQINDTFPMWLGTLPELRVLRLTNSKLHGIRSSGALWEPDLRIIDLSR 754  
 629 --LSTRNPNITSRVYGGHT-----SPTFDN--NGSMM-----FLD 660

Db 755 NAFSODLPTSLFELHKMRVVDKMEPEPSYESYVDSVAVVTGKLEIYRILSLTYIID 814  
 QY 661 MSYNMLSGYIPKEIGSMPLYFLINLGNDSGISPEVEGDLRGINLILDSNKLDRIPQ 720  
 Db 815 LSNKFGHPIPSVLGDLIALRVNLVSHNALQGYIPSSLSGLSILSIESLDSFNQLSGIPQ 874  
 QY 721 AMSALTMTELDLNNLNSGPIPEMGOFETFPAPKFLNBPGLGYPPLPR--C--DPSNADG 777  
 Db 875 QASLTFLEVLNLSHNYLQCTPGQPFRTFESNSYEGNDGLGYPVSKCGCKDVPVSEKN 934  
 QY 778 YA-----HHRSHGR-----BPASLNGSVAMGLLSPFCIFGLIVG-----R 815  
 Db 935 YTVSALDEQESNSRFFNDFWKALMG--YSGGLCIGISIIYILISTGNLWLANIIELEH 993  
 QY 816 EMRRKRRKKEALEMYAEHGNSGDRTANN 845  
 Db 994 KIVQRRKKQ-----RGQRYRRRN 1014

# RESULT 12

US-09-353-585-3  
 Sequence 3, Application US/09353585  
 Patent No. 6287865

## GENERAL INFORMATION:

APPLICANT: Dixon, Mark S  
 Jones, David A  
 Jones, Jonathan DG  
 TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof

NUMBER OF SEQUENCES: 15  
 CORRESPONDENCE ADDRESS:

ADDRESS: Nixon & Vanderhye PC  
 STREET: 8th Floor, 1100 No. 6287865th Glebe Road  
 CITY: Arlington  
 STATE: Virginia

COUNTRY: United States of America  
 ZIP: 22201-4714

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/353,585  
 FILING DATE: 15-Jul-1999

CLASSIFICATION: C12N 15/29, 15/82, A01H 5/00, A01N 65/00, C12O 1/68

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/930,277  
 FILING DATE: 27-Oct-1997

APPLICATION NUMBER: PCT/GB96/00785  
 FILING DATE: 01-Apr-1996

APPLICATION NUMBER: GB 950658.5  
 FILING DATE: 31-Mar-1995

ATTORNEY/AGENT INFORMATION:  
 NAME: Ms Mary J Wilson

REGISTRATION NUMBER: 32,955  
 REFERENCE/DOCKET NUMBER: 620-69

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100  
 INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 1112 amino acids

TYPE: amino acid  
 STRANDEDNESS: <unknown>

MOLECULAR TYPE: linear  
 HYPOTHETICAL: YES

ORIGINAL SOURCE:  
 ORGANISM: Tomato  
 STRAIN: Cf2

SEQUENCE DESCRIPTION: SEQ ID NO: 3  
US-09-353-585-3

Query Match 14.4%; Score 889; DB 4; Length 1112;  
Best Local Similarity 30.6%; Pred. No. 2,6e-56;  
Matches 283; Conservative 129; Mismatches 329; Indels 184; Gaps 30;

```

QY 74 VTSIDLSKPLNNGSAVSSLSLTGLESIFLSNHSVSGFKC-SASLTSLDLSRN 132
DB 217 LTELDSNNALN---GSIPLAGMNNMISFLFLYGNOLSGIPEICYLKSLTYLDLSN 273
QY 133 SLGSPVTLTLTLAGSCSGIKFLNVSNTLDPFGKVSGLK-----LMSLEYLDSANSISG 187
DB 274 ALNDSIP--ASLGNLNNLSFLFLYGNOL-----SGSIPEIYLSLNLVYLGSENALNG 325
QY 188 ANVVGWVLSDCGGLKHLA---ISGNKISGDVDVS--RCVNEFLDVSSNNFSTGIP-FL 241
DB 326 S-----IPASIGMILKMLSRNLNVLNQLSGSIPASIGMNNLMLVYNQLSGSIPASL 379
QY 242 GDCSALQHLDISGNKLSGDFSRASITTEKLANISNQVGPPL--PLKSLQYLSIA 299
DB 380 GNANLMLSLYLYNNQLSGSIPASIGMNNLMLVYNQLSGSIPASIGSIPESIGLSLTYLDLS 439
QY 300 ENKFTGIPD-----FL-----SGACDTLTGLDLSGNHRYGVANPPF 336
DB 440 NNSINGFIPASFGMSMLAFLEYENQLASVPEIGYRSLNVLDSLSENALNGSIPASIF 499
QY 337 GSGGLESLALSNNNSGELPMDTLLKMGKLYDLSPNEFSGELPESTNLASILTLD 396
DB 500 GNANLMLSLYLYNNQLSGSIP-SEIGYRSLNVLDSLSENALNGSIPASIGMNNLMLVYNQLSGSIP 557
QY 397 LSSNFFGFLPLNLCQPKVTLQELVQNNQFTGKLPPLSLNCSSEVLSLHFNYSLTG 456
DB 558 LVNNQLSGSIPPEI--GYRSLNVLDSLSENALNGSIPASIGMNNLMLVYNQLSGSIP 615
QY 457 PSSLSGLSKRLDKLMLNMLEGIPQELMYVKTLETLIDPNDLAGEIPSGLSNCTNLNW 516
DB 616 PEIIGYLSLTYLSLNGNSLNGLIPASFGMNNLMLVYNQLSGSIPESIGLSLTYLDLS 675
QY 517 ISLSNNRLGEIPKMGLENLAILKLSNNNSFGNIPDELGDGCSLIMLDTNLFNGTI 576
DB 676 LYMPRNLKGVQCLGNSLNLQVLSMSNSPSGELPESISMLTSLQTLFORNNLBGAI 735
QY 577 PAAM-----FSGKLANFLAGRYVYKNDGKKE-----CH----- 610
DB 736 PQCFGNISLSEVDMQNNLSGLTPTNFSIGSLISLNLHGNELDEIPSLDNCKQLQV 795
QY 611 -----GAGNLEFQGR--SEOL-----NRLST 631
DB 796 LLDGNDQNLDTFPMWLGTLPELRVLRITSNKLGPIRSSRAIMPPDLRIIDLSRNAFSQ 855
QY 632 RNPNCITSRVYGGHT-----SPTPDN--NGSM-----FLDMSYNN 665
DB 856 DLPTSLFELKGRITVDKTMEEPSYESYDVSVVVTGKLELEIVRLISLTYLDLSNNK 915
QY 666 LSGYLPKEIGSNPYFLIINLGHNDISGIPDEVGDLRGNTLIDSSNNKLDRIQAMASL 725
DB 916 FEGHLPVSLGDLIARILNVSHNALQGITPESIGLSLSELDISFNQLSEIPEQOASL 975
QY 726 TMLTELDSNNLSEPIPMGQFETFPAPKPLNPGLCGYLPR-C--DPSNADGYA--- 779
DB 976 TLELEPLNLSHNYLQGITPQGFRTFESNSYEGNDLGRGVSVSGCGMDPVSEKNYVSA 1035
QY 780 -HHQSHGR-----RPASLGSVAMGLFSPVCFGLIVG-----REMKR 820
DB 1036 LEDQSNSEFPDPMKALMG--YSGGLCTIGISITITLSTGNLMLARIIELEHKIMQ 1094
QY 821 RKKKEALEMTAVGHGNSGDRYANN 845
DB 1095 RKKQO-----RGRNRYRRRN 1110

```

RESULT 13

US-09-180-439-3

Sequence 3, Application US/09180439

Patent No. 6225532

GENERAL INFORMATION:

APPLICANT: Dixon, Mark S

APPLICANT: Hatixanthis, Kostas

APPLICANT: Jones, Jonathan DG

APPLICANT: Jones, David A

TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof

FILE REFERENCE: 620 - 53

CURRENT APPLICATION NUMBER: US/09/180,439

CURRENT FILING DATE: 1998-12-06

EARLIER APPLICATION NUMBER: PCT/GB97/01249

EARLIER FILING DATE: 1997-05-08

EARLIER APPLICATION NUMBER: GB 9609681.3

EARLIER FILING DATE: 1996-05-09

EARLIER APPLICATION NUMBER: GB 9619924.5

EARLIER FILING DATE: 1996-09-24

NUMBER OF SEQ ID NOS: 10

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 3

LENGTH: 968

TYPE: PRT

ORGANISM: Lycopersicon esculentum

US-09-180-439-3

Query Match 14.3%; Score 885; DB 4; Length 968;  
Best Local Similarity 29.6%; Pred. No. 4.2e-56;  
Matches 236; Conservative 152; Mismatches 356; Indels 196; Gaps 37;

```

QY 2 KTFSSFFLVTTLFFSFFSLSPQASPSQSLYREIHOISKVDLPDQ--LTDMSNNK 59
DB 7 KYVSS--LQFFIVFY--LTVAFASST-----EATYFALLTKAFKQNNSSFLASWTSS 56
QY 60 NRC-TEDGYTCDDXYTSDISKPLNVGFSVSSLL-----SLTGLESIFLSNHSNG 113
DB 57 NCKQWYGVVCLNGRVNT-----LNTNASVIGTLVYAFPPSSLPFLNLDLSNNISG 109
QY 114 SV-SGFKCASLTSLSLDRNLSGVTTLTSLGSCSGKFLNVSNTLD--FPGKVSGL 170
DB 110 TTPPELGNLTNLYVDLNTNQLSG--TTPQIGSLAKLQIIRIFNNHNGFIPPEIG-- 164
QY 171 KUNSEVLDLSNLSISGA-----NVVGWVLSDCGGLK-----HLAISG 209
DB 165 YLRSLTKSLGINFLSGSIPASIGMNNLMLVYNQLSGSIPESIGYRSLTKSLD 224
QY 210 NKISGDVVS--RCVNEFLDVSSNNFSTGIP-FLGDCSALQHLDISGNKLSGDFSRAS 266
DB 225 NFLSGSIPASIGMNNLMLVYNQLSGSIPESIGYRSLTKSLGINFLSGSIPASIG 284
QY 267 TCTELKLANISNQVGPPL--PLKSLQYLSAENKFTGEIPDPLSGACDTLTGLDLS 324
DB 285 NLNLSRLDILNNKLSGSIPEIETYLSLTYLDGENALNGSIPSSL--GNLNNLSRLD 343
QY 325 GNHFYGAVPPEFGCSLSELSLSSNNFSGELPMDTLLKRGKLYDLSPNEFSGELPES 384
DB 344 NNKLSGSIPEIETYLSLTYLDGENALNGSIPR--SLGNLNNLMLVYNQLSGSIP 402
QY 365 LTNLSASLTLDSSNNFSGPI--LPLNLCQPKVTLQELVQNNQFTGKIPPLSNGCE 441
DB 403 IGYLS--SLTELTYLNNLSNGSIPASIGNTL-----NNLPMYLYNNQLSGSIP 456
QY 442 LVSLHLSFNYSLTIPSSLSLSTKRLDKLMLNMLEGIPQELMYVKTLETLIDPNDLT 501
DB 457 LTELPLGNNSLNGSIPASIGMNNLMLVYNQLSGSIPASIGMNNLMLVYNQLSGSIP 516
QY 502 GEIPLSGNSCTNLNMLISLNNRLTGEIPKMGLENLAILKLSNNNSFGNIPDELGDGCS 561
DB 517 GEIPLSGNSCTNLNMLISLNNRLTGEIPKMGLENLAILKLSNNNSFGNIPDELGDGCS 576
QY 562 LMLDLNTNLFNGTIP-----AAMF-----KOSGKLAANFLNKGRTVYI----- 600
DB 577 LKILDPGRNNLEGAIPQFGNISLQVFDQNNKLSGLTPTNFSIGSLISLNLHGNELA 636

```

QY 601 -----KNDGMR-----ECHG-----AGNLEF 618  
 Db 637 DEIPSLDNCKKQVLDGNDQNDTFFPMMLGTLPELRVLRILSNLHPISSGAEIMF 696  
 QY 619 QGIRSEOLNR--LSTRNPNITSRYVGGHT-----SPTFDN--NGSNM-----657  
 Db 697 PDRIIDLSRNARSQDLPSTLFEHLKGMRTVDKMEEPSESYSDSVVVVTKGLELEIV 756  
 QY 658 -----FLDMSYNMLSGYIPKEIGSNPYFIINLGHNDISGSIPEVEGDLRGILNIDLS 710  
 Db 757 RIILSYLTIDLSNKKFEGHIPSVLGDLAIRVLNVSNNALOGYIPSSLSGLSILSILSIDS 816  
 QY 711 SNKLDGRIPQMSALTMLTEIDLSNNLSGPIPEMGQFETFPAPKFLNPGLCGYPLPR- 769  
 Db 817 FNQISGEIPQOLASLTLEVLNLSHNYLOGCIPGQPFRTFESNSYEGNDGLGYPVSKG 876  
 QY 770 C--DPSNADGYA--HHQSHGR-----RPASLAGSVAMGLFSPVCIFGLIVG-----814  
 Db 877 CGKDPVSEKNYTVGALDEQESNEFPNDPFWKALMG--YGSGLCIGISIIYILISTGNLRW 935  
 QY 815 -----REMRKRRKKEALEMYAEGHNSGDRYANN 845  
 Db 936 LARIIELEHKIIVORRKKQ-----RGQRNRRRN 966

RESULT 14  
 US-09-180-439-4  
 ; Sequence 4, Application US/09180439  
 ; Patent No. 6225532  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dixon, Mark S  
 ; APPLICANT: Hatzixanthis, Kostas  
 ; APPLICANT: Jones, David A  
 ; APPLICANT: Jones, Jonathan DG  
 ; TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof  
 ; FILE REFERENCE: 620 - 53  
 ; CURRENT APPLICATION NUMBER: US/09/180,439  
 ; EARLIER FILING DATE: 1998-12-06  
 ; EARLIER APPLICATION NUMBER: PCT/GB97/01249  
 ; EARLIER FILING DATE: 1997-05-08  
 ; EARLIER APPLICATION NUMBER: GB 9609681.3  
 ; EARLIER FILING DATE: 1996-05-09  
 ; EARLIER APPLICATION NUMBER: GB 9619924.5  
 ; EARLIER FILING DATE: 1996-09-24  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 4  
 ; LENGTH: 968  
 ; TYPE: PRT  
 ; ORGANISM: Lycopersicon esculentum  
 ; US-09-180-439-4

Query Match 14.3%; Score 884; DB 4; Length 968;  
 Best Local Similarity 29.6%; Pred. No. 5e-56;  
 Matches 296; Conservative 151; Mismatches 357; Indels 196; Gaps 37;  
 QY 2 KTFSSFFLVTTLFFPSFSLSFQASPSQSLYREIHQILSPKDYLPDKA--LLPDMSSNK 59  
 Db 7 KVFSS--LQFTVFFY--LFTVAFASL-----EATATLKKWKATFNQNNNSFLASTYSS 56  
 QY 60 NPC-TFPGVTCRDKYTSIDLSKPLNVGSAVSSSL-----SLTGLSLFSLNSHNG 113  
 Db 57 NACKDWGVVCLNGRVNT-----LNTINASVIGTILYAPFSSLFLENLDSNNNISG 109  
 QY 114 SV-SGFKCSALTSIDLSRNSLSPVTTLSGSCGLKFLNVSNTLD--FPKVSGL 170  
 Db 110 TIPEIGNTLVLDLNTQISG--TIPIQISGLAKLQIRIFNNHNLGFIPEEIG---164  
 QY 171 KNSLEVLIDANSISGA-----NVGWSLSDGGEIK--HLAISG 209  
 Db 165 YLSRLTKLSIGINFLSGSIPASLGNMTNLSFLFLYENQSLGFIPEEIGYLSRLTKLSLDI 224

QY 210 NKISGDVDS--RCVNEFLDVSSNNPSTGIP-FLGDCSALQHLIDSGNKLSGDFSRAS 266  
 Db 225 NPLSGSIPASLGNLNNLSFLYLNQSLGSIPEEIGYLSRLTKLSIGINFLSGSIPASIG 284  
 QY 267 TCTELKTLNINSNQFVPIPL--PLKSLQYLSIAENKFTGEIIPDLFSACDITLTGLDLS 324  
 Db 285 NLNNLSRLDLYNNLSLSSGSIPEEIGYLSRLTYLDGENALNGSIPSSL--GNLNNLSRLDLY 343  
 QY 325 GNHYGAVPPFGSCSLLESALSSNNPFGELPMVDTLKKRGKVLDSLSENESEGEIPES 384  
 Db 344 NNKLGSIPPEEIGYLSRLTYLDGENALNGSIPSA--SGLNANLPMYLYLNQSLGSIPEE 402  
 QY 385 LTNLSASLTLDSSNPGPI--LPNLCONPYTLOEYLONNGFTGKIPTLSNCSB 441  
 Db 403 IGYLS-SLTELXNLNLSNGSIPASLGNL-----NLPMYLYLNQSLGSIPEEIGYLS 456  
 QY 442 LVSLHSPNYLSGTPSSLSGLSKRDLKMLNMBEIGLOELMYVVTLETLIDFENDLT 501  
 Db 457 LTLEFLGNNSINGSIPASLGNLNNLSRLYLNQSLGSIIPASFGNNRNLDTLFLSDNDLI 516  
 QY 502 GRIPSGLSNCTNNLNNLSNRLGEGIPKIGRLNLAALKLSNNSPFGNIPDELGDCRS 561  
 Db 517 GEIPFVCLNLSLEVLNYSRNNLKGVYQCCGNISDLHIILSMSSNSFRGELPSISNLS 576  
 QY 562 LTMDLNLTNLFNGTIP-----AAMF-----KSGKIANFTAGKRYVI-----600  
 Db 577 LKILDPGRNNEGALPQPFNGNISLQVFDQNNKLSTGLPTNFSIGSLSILNLHGNELA 636  
 QY 601 -----KNDGMR-----ECHG-----AGNLEF 618  
 Db 637 DEIPSLDNCKKQVLDGNDQNDTFFPMMLGTLPELRVLRILSNLHPISSGAEIMF 696  
 QY 619 QGIRSEOLNR--LSTRNPNITSRYVGGHT-----SPTFDN--NGSNM-----657  
 Db 697 PDRIIDLSRNARSQDLPSTLFEHLKGMRTVDKMEEPSESYSDSVVVVTKGLELEIV 756  
 QY 658 -----FLDMSYNMLSGYIPKEIGSNPYFIINLGHNDISGSIPEVEGDLRGILNIDLS 710  
 Db 757 RIILSYLTIDLSNKKFEGHIPSVLGDLAIRVLNVSNNALOGYIPSSLSGLSILSILSIDS 816  
 QY 711 SNKLDGRIPQMSALTMLTEIDLSNNLSGPIPEMGQFETFPAPKFLNPGLCGYPLPR- 769  
 Db 817 FNQISGEIPQOLASLTLEVLNLSHNYLOGCIPGQPFRTFESNSYEGNDGLGYPVSKG 876  
 QY 770 C--DPSNADGYA--HHQSHGR-----RPASLAGSVAMGLFSPVCIFGLIVG-----814  
 Db 877 CGKDPVSEKNYTVGALDEQESNEFPNDPFWKALMG--YGSGLCIGISIIYILISTGNLRW 935  
 QY 815 -----REMRKRRKKEALEMYAEGHNSGDRYANN 845  
 Db 936 LARIIELEHKIIVORRKKQ-----RGQRNRRRN 966

RESULT 15  
 US-08-475-891A-2  
 ; Sequence 2, Application US/08475891A  
 ; Patent No. 5859339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ronald, Pamela C.  
 ; APPLICANT: Song, Wen-Yuang  
 ; APPLICANT: Wang, Guo-Liang  
 ; TITLE OF INVENTION: Procedures and Materials for Confering  
 ; TITLE OF INVENTION: Disease Resistance in Plants  
 ; NUMBER OF SEQUENCES: 15  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Townsend and Townsend and Crew LLP  
 ; STREET: Two Embarcadero Center, Eighth Floor  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94111-3834  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,891A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/373,375
FILING DATE: 17-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Baerian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 02370-058910US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1023 amino acids
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Protein
LOCATION: 1..1023
OTHER INFORMATION: /note="Xa21 Xanthomonas spp.
OTHER INFORMATION: disease resistance polypeptide RRR-F
OTHER INFORMATION: from rice (Oryza sativa)"
NAME/KEY: Modified-site
LOCATION: 1010
OTHER INFORMATION: /note="Xaa = Leu when position
OTHER INFORMATION: 5471 of RRR-F = G or Phe when position
OTHER INFORMATION: 5471 of RRR-F = C"
US-08-475-891A-2.

```

```

Query Match 14.2%; Score 875.5; DB 2; Length 1023;
Best Local Similarity 27.6%; Pred. No. 2,3e-55;
Matches 321; Conservative 166; Mismatches 404; Indels 273; Gaps 45;

```

```

QY 89 SAVSSLSITGLSFL--SNSHNGSVGFCASLITSLDLSRNSISG-PVTTLTSLG 145
DB 7 SVMISSLL--LLIGPASDDDAALAAARTSTGVAAATNSCSLSHPCYT--RG 58
QY 146 SCGGKLFVAVSS--TLDPKGYSGGLKINSLEVLDSANSISGANVAVVWGLDGCCE-- 201
DB 59 ACT-LASWNTSGHGHCTWGVVCGARARHPRVVKLLRS--SNLGG-IISPSLGNLS 113
QY 202 -LKHLSISGNKISGDVDSRCVNLRFVDSNNPSTGIPFLGDCSALQHLDISGNKLSGD 260
DB 114 FLRELDLSNYSGEIP--PELSRLSRQLLRLSGNSIQGS 152
QY 261 FSRALSTTELKLNLISSNQFVPIPLK-SLOYLS--LAENKP--TGEI--PDF 310
DB 153 IHAAGACTKLTSLDLSHQ--LRVPAETSLFVPSHQMLCOERPHLLGNLTTPSV 208
QY 311 LSGACDTLT-----GLD--LSGNHFYGAVPFFGSCSLLESIAL--SSNPF 353
DB 209 FDLTCNRLSRSTYFARAQOOSIDYFCATNNLTGMINSIMWLSLAASFCAKRYKLG 268
QY 354 GELPMDTLKMGKLVLLSPNEFSGELPESLTLUSASLTLTLDSSNPFSGPIIPNLCON 413
DB 269 GMIPFNARKTLLHLEVVGMCTRFPGKIPASVANAS--HLTRLQIDGNLFGSIIISGF--G 325
QY 414 PKNLQELYLQNGFTGK-----IPPTLSNCSBLVSIHLSFNLTSGTIPSLGSLSKLR 467
DB 326 RLRLNLTLYLWNLFPQTRQEDMGWIFISDLTNCSTKQTLIDGNNLVGAVLPNPSNLS-- 382
QY 468 DLKMLNMLEGIPQELMYVKTLETLIDFNDLTGEIIPSGLSNCTNLNMIISLNNRLTGE 527
DB 383 -----TSLSFLALDLNKITGSIPIKDIGNLIGLHLYLCNNPFRGS 422

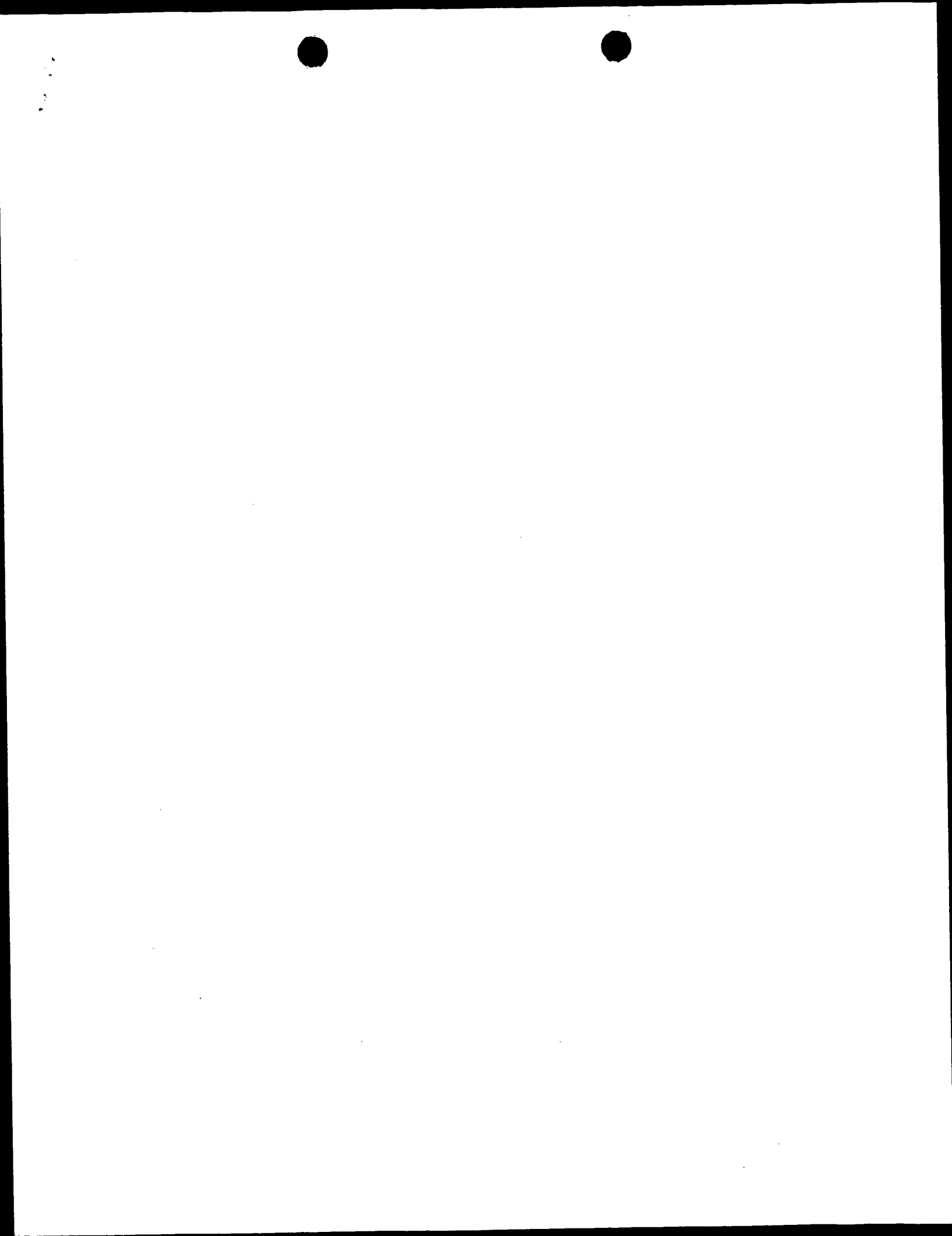
```

```

QY 528 IPKNIIGLEMLAILKLSNNSPSGNIPDELDCRSLIMLDLNTLFGFTLPAMFKOSGI 587
DB 423 LPSSIGRLRNGLIVAVENNLSSIPALQNLLEMLNLLGNTKBSGWLPTLSNLTNLL 482
QY 588 AANFIAGKRYVYIYNDGKKCKGAGNLEFQGISSEQLNRSLSTNPNCTISRYVGHTS 647
DB 483 SL-----GLSHLAPQSGLDL--YTSQPNHCIVATPISGSGOI 517
QY 648 PTFD-----NNGSMFLDMSYNNLSGY-IPKETSMPYFIATLGHNDISGIPDEVDL 701
DB 518 PQVTLQIVQTPPKKMINVSKNTLGGIRYPQETGHNKLVEMQNRIDIS-KIPVTLGDC 576
QY 702 RGNILDLSSNKLDRIPQAMSLTMTLEIDLSNNNLSPGIPEMQGFEPFPAKFLNN-- 759
DB 577 QLRVLYLQNNLSSGIPALGQKLEFLDLSNNLSQITP--SLSRYYASFLEPFF 634
QY 760 PGLCGYPLPRCDPSNADGYAHH--QRSHGRPPASLAGSVAMGLFSGVCI FGLILVGR 815
DB 635 QQLCG-----GSANHCVLSSQMHGSGSKAMPNSVVEYLI--YICLDVPELL-- 678
QY 816 EMRK-----RRKKAELMVAEGHNSGSDRTAANTN 847
DB 679 ENRKHFPALPISVLAALAILSSLYLLITWNKRYKGAAPRTSKWGH----- 726
QY 848 WKLTGVKALSLINLAFFKPLRKLTFADILQATNGFNHDSLIISGSGFGDYKAILKXGSA 907
DB 727 -----PL--VSYPOLVYATDGFAPTNLGSGSFASVYKRLLENPKA 765
QY 908 VAIKLIHVSQGDREPMAMETIKIRKRNLPVLLGYCKYGDER-----LVNENVMKYG 962
DB 766 L-----KSFPAECBALNMRHRLVKLVITCSSIDNRGNDFKAIYVDFMPNG 812
QY 963 SLSDVLA--DPKKGVTLKSTRKKAIGSARGIAFLHNCSPHIIRDMKSSNVLLDE 1019
DB 813 SLEDWHPETNDQDORHLNCHRRVLTILLDVAICALDYLRHGRPEPVVHCDVKSNNVLLDS 872
QY 1020 NLEARVSDFGMARLM-----SAMDTLSVSTLAGTGPVYVPEYQSFRCSTKGDVYSGVV 1075
DB 873 DMVAHVGDGSLARILVDGSLIQOSTSMGFRGTTGYAABPYGVGHIASTHGDIYSYGL 932
QY 1076 LLELTLGRPTDS--PDFGNNLVGWVKQHAKL-----RISDVDPPELM-----K 1118
DB 933 VLEIYVGRKPTDSTRPDLG-----LRQYVELGLHGRVTDVDTLLIDSEWNLNSTN 985
QY 1119 EDPALER--LLQHLKAVACLDD 1140
DB 986 NSPCRRTCEIVSLRLGLSCSD 1009

```

Search completed: March 10, 2003, 18:19:21  
Job time : 33.6693 secs





new receptor kinase gene involved in brassinolide signalling useful for promoting increased yield and disease resistance in

plants and for modulating oocyte maturation  
claim 3; Page 52; 72pp; English.

This is the amino acid sequence of a novel plant steroid receptor kinase, designated Bin1, which is involved in the pathway for the synthesis of the plant steroid hormone, brassinolide. 18 New Arabidopsis dwarf mutants were identified that lacked the ability to respond to brassinolide, and were named bin mutants. The bin1 mutations were used to map the gene to a small interval on Arabidopsis chromosome 4. Bin1 was cloned by map-based cloning. A Bin1 polynucleotide (see AAX07356) was identified within this interval by sequencing the wild-type and mutant alleles of this nuclear acid. Overexpression of Bin1 in transgenic plants provides plants characterized as having enhanced biomass and increased seed yield. Bin1 expression may also increase resistance to pesticides. Inhibition of Bin1, e.g. using antisense oligonucleotides (AON), is used to render plants male-sterile, and to reduce their stature or yield, e.g. for creating dwarf varieties. Since Bin1 homologues may be involved in regulation of the menstrual cycle and uterine function, Bin1, antibodies and AON may be useful as contraceptives, for improving success of in vitro fertilisation and to prevent premature labour. Transgenic animals are also provided, and are models for studying steroid-receptor interactions or can be used to screen for therapeutic agents.

Sequence 1196 AA:

Query Match 100.0%; Score 6183; DB 20; Length 1196;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MKTSSFFLSATTTLPFFSFFSLSFQASPSQSLYREIHOLISFKVLPDKXLLPDMSSNKN 60  
1 MKTSSFFLSATTTLPFFSFFSLSFQASPSQSLYREIHOLISFKVLPDKXLLPDMSSNKN 60  
1 PCTPDGVTCTRDDKVTISLSSKPLNVSFAVSSSLSTGLESFLNSHINSVSGKCC 120  
61 PCTPDGVTCTRDDKVTISLSSKPLNVSFAVSSSLSTGLESFLNSHINSVSGKCC 120  
61 PCTPDGVTCTRDDKVTISLSSKPLNVSFAVSSSLSTGLESFLNSHINSVSGKCC 120  
121 SASLTSLSLRSNLSGPTVTLTSLGSCGSLKPLNVSNTLPDGVSGGLKLNLSLEVDL 180  
121 SASLTSLSLRSNLSGPTVTLTSLGSCGSLKPLNVSNTLPDGVSGGLKLNLSLEVDL 180  
121 SASLTSLSLRSNLSGPTVTLTSLGSCGSLKPLNVSNTLPDGVSGGLKLNLSLEVDL 180  
181 SANSISGANVGVWVLSDCGGLKHLAISGNKISGDVDSRCVNLKFLDVSSNNFSTGIPF 240  
181 SANSISGANVGVWVLSDCGGLKHLAISGNKISGDVDSRCVNLKFLDVSSNNFSTGIPF 240  
181 SANSISGANVGVWVLSDCGGLKHLAISGNKISGDVDSRCVNLKFLDVSSNNFSTGIPF 240  
241 LQDCSALQHLDISGNKLSGDFSRALSTCTELKLNISNQFPGPIPLPKLSLOYLSLAE 300  
241 LQDCSALQHLDISGNKLSGDFSRALSTCTELKLNISNQFPGPIPLPKLSLOYLSLAE 300  
301 NKFTGEIPPLFGACDTLTGLDLSGNHFGAVPPFGSCSLLESIALSSNNFSGELPMDT 360  
301 NKFTGEIPPLFGACDTLTGLDLSGNHFGAVPPFGSCSLLESIALSSNNFSGELPMDT 360  
361 LLMRGLKVLDSFNFSGELPESLTNLSASLTITDLSNNFSGPIPLPMLCONPKXITLOE 420  
361 LLMRGLKVLDSFNFSGELPESLTNLSASLTITDLSNNFSGPIPLPMLCONPKXITLOE 420  
421 LVIQNNFGTKIPPTLSNCSGLVSLHSFNYSGLTIPSSLSGSLSKRDKLKMLNMLEGRI 480  
421 LVIQNNFGTKIPPTLSNCSGLVSLHSFNYSGLTIPSSLSGSLSKRDKLKMLNMLEGRI 480  
481 POELMTVKTLETLIDFNLDLGEIPGLSNCTVNMWISLNNRLTGEIPKWIQRLENTLAI 540  
481 POELMTVKTLETLIDFNLDLGEIPGLSNCTVNMWISLNNRLTGEIPKWIQRLENTLAI 540  
541 LKLSNNSFSGNIPDELGDCLSLIWLDTNLTENGITPAAMFKOSGKIAPNAGKRYVYI 600  
541 LKLSNNSFSGNIPDELGDCLSLIWLDTNLTENGITPAAMFKOSGKIAPNAGKRYVYI 600  
541 LKLSNNSFSGNIPDELGDCLSLIWLDTNLTENGITPAAMFKOSGKIAPNAGKRYVYI 600

601 KNDGKKCHGAGNLEFFQGRBSQNLRLSTRNPNITSRVYGHTSPFPDNNGNMFLD 660  
601 KNDGKKCHGAGNLEFFQGRBSQNLRLSTRNPNITSRVYGHTSPFPDNNGNMFLD 660  
661 MSYMLSGYIPKEIGSMPLYFIINLGHNDISGIPDEVDLRLGILNLDSSNKLDGRIPQ 720  
661 MSYMLSGYIPKEIGSMPLYFIINLGHNDISGIPDEVDLRLGILNLDSSNKLDGRIPQ 720  
721 AMSALVMTLEIDLSNNLSGPIPMKQGFETPPPAKFLANNGLGYPILPRCDPSNADGYAH 780  
721 AMSALVMTLEIDLSNNLSGPIPMKQGFETPPPAKFLANNGLGYPILPRCDPSNADGYAH 780  
781 HORSGRBPASLAVMAGLIFSVCIFGLILVGRERKRRKKEALEMYABEGHNSGD 840  
781 HORSGRBPASLAVMAGLIFSVCIFGLILVGRERKRRKKEALEMYABEGHNSGD 840  
841 RTANNTNKLTVGEKALSLINLAFEKPLRKLTPADLLQATNGFHNDSLISGGGVDYKA 900  
841 RTANNTNKLTVGEKALSLINLAFEKPLRKLTPADLLQATNGFHNDSLISGGGVDYKA 900  
901 IIKQSSAVAIAIKKLIVHSGQGRBEMAEETIGIKHNLVPLGLCYCKVGRDLVNEVMK 960  
901 IIKQSSAVAIAIKKLIVHSGQGRBEMAEETIGIKHNLVPLGLCYCKVGRDLVNEVMK 960  
961 YGSLBDVLQDPKKGAVKJLSTRKRIAGSARGLAFHNCSPHIIHRDKSNVLLDEN 1020  
961 YGSLBDVLQDPKKGAVKJLSTRKRIAGSARGLAFHNCSPHIIHRDKSNVLLDEN 1020  
1021 LBARVSDFGMARLMSAMDTLHLSVSTLAGTPGVVPEYVQSRFSTKGDVYGVVLELL 1080  
1021 LBARVSDFGMARLMSAMDTLHLSVSTLAGTPGVVPEYVQSRFSTKGDVYGVVLELL 1080  
1081 TKRRPDSPPDGNLNVGVWQKALRISDVDPPELMKEDPALTEILLQHKVAVACLD 1140  
1081 TKRRPDSPPDGNLNVGVWQKALRISDVDPPELMKEDPALTEILLQHKVAVACLD 1140  
1141 RAMRPTVQVWAMFEKIOAGSIDSGTIRSIEDGFSITIBMDMSIKVPEKGL 1196  
1141 RAMRPTVQVWAMFEKIOAGSIDSGTIRSIEDGFSITIBMDMSIKVPEKGL 1196

RESULT 2  
AAB67443  
ID AAB67443 standard; Protein; 1196 AA.  
XX  
AC AAB67443;  
XX  
DT 15-MAY-2001 (first entry)  
XX  
DE Amino acid sequence of an Arabidopsis Br1 protein.  
XX  
KW Disease resistance protein; Xa21; RKK gene; transgenic plant;  
XX Xanthomonas; plant pathogen; Br1 protein; RCH10 protein.  
XX  
OS Arabidopsis sp.  
XX  
PN MO200109283-A2.  
XX  
PD 08-FEB-2001.  
XX  
PF 28-JUL-2000; 2000MO-US20714.  
XX  
PR 28-JUL-1999; 99US-036313.  
XX  
PA (REGC) UNIV CALIFORNIA.  
XX (SHUK) SALK INST BIOLOGICAL STUDIES.  
XX Ronald P. He Z, Chory J, Lamb C, Li J;  
XX WPI; 2001-159858/16.  
XX N-PSDB; AAF54982.  
XX  
XX Chimeric plant receptors comprising a polynucleotide encoding a RKK

receptor containing a heterologous extracellular domain and a kinase domain from a Xa21 polypeptide, useful for modulating plant responses to pathogens

Disclosure; Page 38-42; 47pp; English.

The present sequence represents a Br1 protein. The specification describes chimeric receptors which are used for modulating plant responses to pathogens. The receptors comprise a heterologous extracellular domain (e.g. from a Br1 protein or Rch10 protein) and a kinase domain (e.g. from disease resistance protein Xa21). The Xa21 gene is a member of disease resistance genes referred to as RKK genes. When Xa21 is present in a transgenic plant, it confers resistance to Xanthomonas spp.. The plant receptors containing heterologous domains are useful for modulating plant responses to pathogens including viruses, bacteria, nematodes, fungi or insects. The nucleic acids can be used to confer desired traits on essentially any plant.

Sequence 1196 AA;

Best Match 99.3%; Score 6140; DB 22; Length 1196;

Best Local Similarity 99.4%; Pred. No. 0; Matches 1199; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

QY 1 MKTFFSFFLSVTLTLPFSEFSLSFQSPSSQLYREIHQILSFQVLPDKMLPDMSSNKK 60
DB 1 MKTFFSFFLSVTLTLPFSEFSLSFQSPSSQLYREIHQILSFQVLPDKMLPDMSSNKK 60
QY 61 PCTFDVTCDDKVTSLDSSKPLNVGFSAVSSSLSLTGLTSLFLSNHINGSVGFPC 120
DB 61 PCTFDVTCDDKVTSLDSSKPLNVGFSAVSSSLSLTGLTSLFLSNHINGSVGFPC 120
QY 121 SABLTSLSLDRNLSGCVTTTLTSLGSCSGLKPLNVSSNTLDPFGKXSGGLKNSLEVL 180
DB 121 SABLTSLSLDRNLSGCVTTTLTSLGSCSGLKPLNVSSNTLDPFGKXSGGLKNSLEVL 180
QY 181 SANSISGANVGVNVLSDGCGELKHLAISGKISGVDVSRCVNLEFLDVSSNNFSGTIRP 240
DB 181 SANSISGANVGVNVLSDGCGELKHLAISGKISGVDVSRCVNLEFLDVSSNNFSGTIRP 240
QY 241 LGDCSALQHLIDISGNLSDGFSRAISTCTELKLNISNOFVPIPLPKSLQVYSIAE 300
DB 241 LGDCSALQHLIDISGNLSDGFSRAISTCTELKLNISNOFVPIPLPKSLQVYSIAE 300
QY 301 NKRTGIIIPPLSACCTITLGLDSGNHFGAVPPFGSCSLLESLASLSSNNFSGELPMDT 360
DB 301 NKRTGIIIPPLSACCTITLGLDSGNHFGAVPPFGSCSLLESLASLSSNNFSGELPMDT 360
QY 361 LKMRGLKYLDSFNFSEGLPESLTNLSASILLTLDLSSNNFSGPTLPMVLCNPKNTIOE 420
DB 361 LKMRGLKYLDSFNFSEGLPESLTNLSASILLTLDLSSNNFSGPTLPMVLCNPKNTIOE 420
QY 421 LYLQNNGFTGKIPTLSSNCELSVSLHSFNYSGLTIPSSISGLSKLRLDKMLNLEGEI 480
DB 421 LYLQNNGFTGKIPTLSSNCELSVSLHSFNYSGLTIPSSISGLSKLRLDKMLNLEGEI 480
QY 481 POELMAYVTKLETLIDFNLDLGEIPSGLSNCTNLNWTISLNNRLTGEIIPKWIIGLENTAI 540
DB 481 POELMAYVTKLETLIDFNLDLGEIPSGLSNCTNLNWTISLNNRLTGEIIPKWIIGLENTAI 540
QY 541 LKLSNNSFSGNTPDELIGDRSLIWLIDLNTNLNFGTIPAMFQSGKIANFLAGRRYYI 600
DB 541 LKLSNNSFSGNTPDELIGDRSLIWLIDLNTNLNFGTIPAMFQSGKIANFLAGRRYYI 600
QY 601 KNDGKMKKECHGAGNLEFQIRSEQLNRLSTNPNCTISRYVGGHTSPFDNNGSMFLD 660
DB 601 KNDGKMKKECHGAGNLEFQIRSEQLNRLSTNPNCTISRYVGGHTSPFDNNGSMFLD 660
QY 661 MSYNMLSGYIPKISGMPYLFILNIGHNDISGIPDEVGDLAGNITLDSNKKLDGRIPQ 720
DB 661 MSYNMLSGYIPKISGMPYLFILNIGHNDISGIPDEVGDLAGNITLDSNKKLDGRIPQ 720
QY 721 AMSALTMLTEIDLSSNNLSGPIPEMGQETFPFAKFLNPGLCGYPLPCDPSNADGVAH 780

```

```

DB 721 AMSALTMLTEIDLSSNNLSGPIPEMGQETFPFAKFLNPGLCGYPLPCDPSNADGVAH 780
QY 781 HQRSHGRRPASIASVAMGLLSPVCFIFGLILVGRMKRRRKRKBALEMYAEGHNSGD 840
DB 781 HQRSHGRRPASIASVAMGLLSPVCFIFGLILVGRMKRRRKRKBALEMYAEGHNSGD 840
QY 841 FRANNNTMKLVGVKALSLINLAFFKPKRLTPADLQATNGFNHDSLISGGGSDVYKA 900
DB 841 FRANNNTMKLVGVKALSLINLAFFKPKRLTPADLQATNGFNHDSLISGGGSDVYKA 900
QY 901 ILKDSAAVAIKLILVSGQDREFMAEMETIGKIKHRLVPLIGYCKVGDRLLVNEMK 960
DB 901 ILKDSAAVAIKLILVSGQDREFMAEMETIGKIKHRLVPLIGYCKVGDRLLVNEMK 960
QY 961 YGSLDEVQDPKKGAVKALKSTRKRIAGARGIAPLHNCSPHIIHRDKSSNVLLDEN 1020
DB 961 YGSLDEVQDPKKGAVKALKSTRKRIAGARGIAPLHNCSPHIIHRDKSSNVLLDEN 1020
QY 1021 LEARVSDFGMARLMSAMDTLHVSSTLACTPGVPEYVOSFRCTGDIVSYGVVLELL 1080
DB 1021 LEARVSDFGMARLMSAMDTLHVSSTLACTPGVPEYVOSFRCTGDIVSYGVVLELL 1080
QY 1081 TGRKPTDSDPDGNNLVGVVQKAKRLISDVPELMEKEDPALEIELLOHLKAVACLDD 1140
DB 1081 TGRKPTDSDPDGNNLVGVVQKAKRLISDVPELMEKEDPALEIELLOHLKAVACLDD 1140
QY 1141 RAMRPTVQVMAFKEIQAGSGIDSGTIRSIDGPGSTIEMVMSIKEVPEGL 1196
DB 1141 RAMRPTVQVMAFKEIQAGSGIDSGTIRSIDGPGSTIEMVMSIKEVPEGL 1196

RESULT 3
AAE19490
ID AAE19490 standard; Protein; 1196 AA.
AC AAE19490;
AD 31-MAY-2002 (first entry)
DE Brassinosteroid receptor protein encoded by Br1 DNA.
DS Plant receptor protein; disease resistance protein; plant development;
KW RRR protein; disease resistance; cell free assay; gene therapy;
KM br1 DNA; brassinosteroid protein.
XX OS
XX Unidentified.
FH Key Location/Qualifiers
FT Misc-difference 162
FT /note= "Encoded by TTT"
XX
XX MO200210367-A1.
XX
XX 07-FEB-2002.
XX
XX 28-JUL-2000; 2000MO-US20604.
XX
XX 28-JUL-2000; 2000MO-US20604.
XX
XX (SALK ) SALK INST BIOLOGICAL STUDIES.
XX
XX Chory J, Lamb C, He Z;
XX
XX WPI; 2002-227084/28.
XX
XX N-PSDB; AAD30800.
XX
XX Heterologous polynucleotide encoding chimeric plant receptors for
XX controlling plant development and disease resistance, has leucine-rich
XX repeat domain, transmembrane domain, and cytoplasmic protein kinase
XX domain
XX
XX Disclosure; Page 49-50; 54pp; English.

```

XX The present invention relates to heterologous nucleic acid molecules  
 CC encoding chimeric plant receptor proteins comprising a leucine-rich  
 CC repeat (LRR) domain, a transmembrane domain and a kinase domain from  
 CC cytoplasmic RRR (disease resistance proteins) proteins such as Xa21.  
 CC The nucleic acid sequences are useful for identifying ligands for  
 CC receptor or receptor-like kinase. The chimeric receptors are useful  
 CC for controlling plant development and/or disease resistance. They are  
 CC used in cell free assay useful for determining the ability of a test  
 CC compound to bind to or modulate the activity or expression of the  
 CC receptor. Sequences of the invention are also used in gene therapy.  
 CC The present sequence is brassinosteroid receptor protein encoded by  
 CC brr1 DNA. This sequence is used in the invention.

XX Sequence 1196 AA;

Query Match 99.1%; Score 6130; DB 23; Length 1196;  
 Best Local Similarity 99.3%; Pred. No. 0;  
 Matches 1188; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

1 MKTFSSFLSVTLTFFSFSLFQASPSQSLYREIHQILFQKVLDPKLLPWSNKN 60  
 1 MKTFSSFLSVTLTFFSFSLFQASPSQSLYREIHQILFQKVLDPKLLPWSNKN 60  
 61 PCTPGVTCRDDKTSIDLSKPLNMGSAVSSSLTGLBSLFLSNHSHNGSVGFKC 120  
 61 PCTPGVTCRDDKTSIDLSKPLNMGSAVSSSLTGLBSLFLSNHSHNGSVGFKC 120  
 121 SASLISLDSRSLSGPYTTLTSLGSCGKLKPLNVSNTLDFPGKVSGLKLNLSLEVL 180  
 121 SASLISLDSRSLSGPYTTLTSLGSCGKLKPLNVSNTLDFPGKVSGLKLNLSLEVL 180  
 181 SANSISGANVGVWVSDGCGELKHLAISGNKISGVDVSRCNLEFLDVSSNNSTGIPF 240  
 181 SANSISGANVGVWVSDGCGELKHLAISGNKISGVDVSRCNLEFLDVSSNNSTGIPF 240  
 241 LGDCSALQHLIDISGNKISGDFSRATSTCTELKLNISNQFGPIPLPLKSLQTLIAE 300  
 241 LGDCSALQHLIDISGNKISGDFSRATSTCTELKLNISNQFGPIPLPLKSLQTLIAE 300  
 301 NKFTGEIPDFLSGACDTLGLDLSGNHFGAVPPFGSCSLLESIALSSNNFSGELPMDT 360  
 301 NKFTGEIPDFLSGACDTLGLDLSGNHFGAVPPFGSCSLLESIALSSNNFSGELPMDT 360  
 361 LKMRGILKVLDSFNFSGELPESLTNLASLITLDSNNNSGPIPLPMLCONPKTLOE 420  
 361 LKMRGILKVLDSFNFSGELPESLTNLASLITLDSNNNSGPIPLPMLCONPKTLOE 420  
 421 LKLNNGGFTGKIPPTLSNCSSELVSLHSFNYSGLTIPSSLSGLSKLRDLKMLNMBE 480  
 421 LKLNNGGFTGKIPPTLSNCSSELVSLHSFNYSGLTIPSSLSGLSKLRDLKMLNMBE 480  
 481 POELMTVKTLETLIDPNDLTGEIPSGLSNCTNLMNLSNNRLEGEIPKIGLENLAI 540  
 481 POELMTVKTLETLIDPNDLTGEIPSGLSNCTNLMNLSNNRLEGEIPKIGLENLAI 540  
 541 LKLSNNSPSENTIDELGDCRSILWLDLNTNLNNGTIPAMFQSGKIANPAGKRYVI 600  
 541 LKLSNNSPSENTIDELGDCRSILWLDLNTNLNNGTIPAMFQSGKIANPAGKRYVI 600  
 601 KNDGKKECHGAGNLLEFQGISREQLNRLSTRNPNCTISRYVGHTSPFPDNGSMFLD 660  
 601 KNDGKKECHGAGNLLEFQGISREQLNRLSTRNPNCTISRYVGHTSPFPDNGSMFLD 660  
 661 MSYNMLSGYIPKEIGSMPIYFIINLGNDISGSIPEVGDRLGINTLIDSSNKLDRIPQ 720  
 661 MSYNMLSGYIPKEIGSMPIYFIINLGNDISGSIPEVGDRLGINTLIDSSNKLDRIPQ 720  
 721 AMALMTLTLTIDLSNNNLSGPIDEMGOFFEPAPKFLANNGLGYPPLPCDPENADGYAH 780  
 721 AMALMTLTLTIDLSNNNLSGPIDEMGOFFEPAPKFLANNGLGYPPLPCDPENADGYAH 780  
 781 HORSHGRPPASLAGVAMGLLFSFVCI FGLILVGREMRKRRKKALEMYAEHGNSGD 840

DB 781 HORSHGRPPASLAGVAMGLLFSFVCI FGLILVGREMRKRRKKALEMYAEHGNSGD 840  
 QY 841 RTANNTMKLTGVKEALSTINLAEPKRLKLTENDLLOANTGFENDLSIGSGGVYKA 900  
 DB 841 RTANNTMKLTGVKEALSTINLAEPKRLKLTENDLLOANTGFENDLSIGSGGVYKA 900  
 QY 901 ILKDSAVAIAIKKLHVSGQGRFPAEMETIGKIKHNLVPLGYCKVGERLLVNEVMK 960  
 DB 901 ILKDSAVAIAIKKLHVSGQGRFPAEMETIGKIKHNLVPLGYCKVGERLLVNEVMK 960  
 QY 961 YGSLDVLQDPKKGKVKLSTTRKRIAGSAGLAFLHNCSPHIIHRDMKSSVLLDEN 1020  
 DB 961 YGSLDVLQDPKKGKVKLSTTRKRIAGSAGLAFLHNCSPHIIHRDMKSSVLLDEN 1020  
 QY 1021 LBARVSDGEMALMANTHLSYSTLAQTPGVPEYQSFRCSTKGVYGVVLELL 1080  
 DB 1021 LBARVSDGEMALMANTHLSYSTLAQTPGVPEYQSFRCSTKGVYGVVLELL 1080  
 QY 1081 TGRPTDSDPDGNNLVGVVQKAKLRISDVDPPELMKEDPALIEILLQHLKVAACIDD 1140  
 DB 1081 TGRPTDSDPDGNNLVGVVQKAKLRISDVDPPELMKEDPALIEILLQHLKVAACIDD 1140  
 QY 1141 RWRREPTVQVAMFEKIOAGSGIDSGTINSIEDGFTIEMVDMSTKEYPEGKL 1196  
 DB 1141 RWRREPTVQVAMFEKIOAGSGIDSGTINSIEDGFTIEMVDMSTKEYPEGKL 1196

RESULT 4  
 AAG79244  
 ID AAG79244 standard; Protein; 1121 AA.

AC AAG79244;  
 XX  
 DT 03-JAN-2002 (first entry)

DE Amino acid sequence of a gene concerning brassinosteroid sensitivity.  
 XX  
 KW OGBR1; brassinosteroid sensitivity; d61 locus; rice;  
 XX  
 OS internode elongation; internode cell; lamina joint.

OS Oryza sativa.  
 XX  
 PN WO200173036-A1.  
 XX  
 PD 04-OCT-2001.

PF 30-MAR-2001; 2001WO-UP02770.  
 XX  
 PR 31-MAR-2000; 2000JP-0101276.

PA (NAG-) NAT INST AGROBIOLOGICAL SCI.

PI Tanaka H, Kayano T, Matsunaka M;  
 XX  
 PT WPI; 2001-616505/71.  
 XX  
 PS N-PSDB; AAI65842.

PT Gene relating to brassinosteroid-sensitivity of plants, useful in  
 CC controlling growth and development of transformants including rice to  
 CC improve harvest and crop yield for animal feed or dwarfism to enhance  
 CC ornamental effect -

PS Claim 1; Page 72-80; 87pp; Japanese.

XX The present sequence is encoded by a gene, designated OGBR1, which  
 CC enhances the brassinosteroid sensitivity of plants. The OGBR1 gene is  
 CC located to the d61 locus. The gene is involved in the growth and  
 CC development of rice, for example, internode elongation via the induction  
 CC of the elongation of internode cells and bending of lamina joints.  
 CC The OGBR1 gene is useful in controlling growth and development of  
 CC transformants. Transformant rice plants can be obtained to improve  
 CC harvest and crop yield for providing more animal feed in agriculture,

CC or plants with dwarfism can be produced to enhance ornamental effect  
CC and added value in horticulture and related industries.  
XX

Sequence 1121 AA;

Query Match	49.3%	Score	3050.5;	DB	22;	Length	1121;
Best Local Similarity	53.9%	Pred	No	6	9a-323.		

Matches 632; Conservative 166; Mismatches 276; Indels 99; Gaps 20;

Qy	39	LISFKVLDKXNLLPMSNKNRPTCPDGYTCDDKYTISIDUSSKPLNVGSVSSLLST	98
Db	31	LEEFRAVPAQAAALKWMSGGDGACFPFGAGCNGRLTSLSLAGVPLNAFPAVATILQ	90
Qy	99	TGLESLEFSNHSNGVS---GFKSASLTSLDLSRN-SLSGPYTLTSLGS-CSGKFL	133
Db	91	GSVEVLSLGAIVSGALSLAAGARGSKLQALDLSGNALGSAVDAAALASAGGKLT	150
Qy	154	NYSNLTLPFGKY--SGKJNSLEVLDSLANSISGANVGVWSDCGGELKHLAISGN	211
Db	151	NLSGDVVG-AAKYGGGGGGGFPAGLDSLDSNNKIIPDSDLAWMDAGAVRWLDLNR	209
Qy	212	ISGDVVRGCVNLEFLDVSNNPFGICIFPLDGCALQHLDISGNKLGDF-SRAISTCE	270
Db	210	IS-----GVEPFTNCSGLQYLDLSGNLIVGEVPGGALSDCG	246
Qy	271	LKLINISSNOFVPIPLPLKSLQYLSLAENKFGELIPDLISGACDTLTGLDLSGNHFG	330
Db	247	LKVTN-----ISFNHLAG	259
Qy	331	AVPPFGCSLLESLTALSSNNPFGELPMDTLKXRGKVLDSLFSNEPFGELPESLTWLSA	390
Db	260	VEPPDIAGLTSLNALTNNPFGELPGEAPAKLQOULTALSLSPHNHNGSIPPTVASL-P	318
Qy	391	SLTLTDLSSNNSGPIILNLTQNPNTLQELYLQNGFTGKIPTLSSNCELYSLHSEN	450
Db	319	ELQOQDLDSNNTSGTIPSLTQDPNPSKXHLTYLQNNYLTGSIIPDVAISCTSLVSLDLSN	378
Qy	451	YLSGTIPSSLSGLSTRLBKLTMLNMLEGEPQELMYKTELTLLPNDLTGSIIPSGLSN	510
Db	379	YINGSIIPASLDGLNLTOLIMQNELBEIIPASLSRIGLEHLILDNGTSGIIPELAK	438
Qy	511	CYNLWISLNNRLTGEIIPKWTIGRIENLIALIKLSSNSPSGNIPELDGCSLITWLDNTN	570
Db	439	CTKXNWLISLNNRSLSGPIPSWLGKLSYALIKLSSNSFSGPIPELDGQSLWLDLNSN	498
Qy	571	LENGTIPAMKSGSKLAANFIACKRVYIKNDCKKECHGAGNLEFQGISBQNLRLS	630
Db	499	QLNGSIPKELKQSGKXNVLIVRPYVYIRNDELSECHGKSLLEFPIRDDLSPRP	558
Qy	631	TRNPCTNSRYVGGHTSPTPDNNSNMVFLDSYNNLSGYPKFEGISNPYLFILNIGHNI	690
Db	559	SKKLCSNFT-RMYVGSTETTFNKGSMIFDLSTNQDLSAIPGELGDMFIMLTNLSHNL	617
Qy	691	SGSIPDEVGDLRGNIILDLSNKLDSRIFQAMSAITWLTEDLSSNNLSGPIPEMGQFET	750
Db	618	SGTIPSLAEAKKLAVIDLSYNQLEGPIMSPFALS-LSBINLSSNQNGTIPELGSLAT	676
Qy	751	PPPKFLNPNPGLCYPIPCPCPSNADQYAHQKSHRRRPRASLAGSYAMGLLSPFCISGL	810
Db	677	FPKQYERNYGLCGFPIPPCDHSPRSSNDHQ-SH-RQASMASSIAMGLLPSLFCIIVT	734
Qy	811	ILVREWRKRRKKAEVL--EMTAEHGNSGDRTRANTYMK--LTGVKALINILAAPEK	866
Db	735	IIAIGSRKRRRLKNEASTSDIYIDSRHS---ATMNSDPRKQLSST-NLSINILAAPEK	790
Qy	867	PLRKLTPADILLQANGHNDSLGSFGFGYVYAILKDSGAVALIKKLHVSGQDREFMA	926
Db	791	PLQKLTADLVEANGPHIALCOIGSGFGYVYKQKDKGVAAIKKLHVSGQDREFTA	850
Qy	927	EMETIGIKIRHNVPLLYGCKVDRLVNEWKYGSLEDTYODPKGVGTXLSTREKI	986
Db	851	EMETIGIKIRHNVPLLYGCKAGEERLIYDYMKFSSLEDTVLDHDKTKLQKXNWEARRKI	910

QY	387	AISGARGLAFLHNCSPHIIHRMKSNNVLBENEARSDGMRILMSADTHLSVSTL	10466
Db	911	AVGARGLAFLHNCPLIIHRMKSNNVLDEQLEARPSDFGMARILMSVVDTHLSVSTL	970
QY	1047	AGTGPVPEPYOSFRCSTKGDVYSYGVALLBELTNGKRPDSDPDG--DNLLVGWYKHAH	11059
Db	971	AGTGPVPEPYOSFRCCTKGDVYSYGVALLBELTNGKRPDSDADGEDNNLVGWYKHTK	10300
QY	1106	LRTSDVDEPLMEDPALLETLOHLKAVAVCLDPRAMRPTVMVMAMFKEIQAGSID	11655
Db	1031	LKTTDVFDEPLLEKEDSVLELLEHAKIACACLDIDRPSRRPMLKVMAMFKEIQAGSTVD	10900
QY	1166	SOST---IRSIEDGSGSTIEWYMSIKVEPBEK	1195
Db	1091	SKTSSAAGSIDGSGY---VLDMPLEKBEK	1120

RESULT 5  
ABB92294

AC ABB92294 ;  
VV

DT 31-MAY-2002 (first entry)  
XX

Herbicide active polypeptide SEQ ID NO 1505.

Herbicide; plant; agriculture; herbicide.

OS Arabidopsis thaliana.  
XX

PN WO200210210-A2  
XX

PD 07-FEB-2002.

PF 28-AUG-2001; 2001WO-EP09892.  
YY

PR 28-AUG-2001; 2001WO-EP09892.  
YY

PA (FARB ) BAYER AG.  
VV

Tietjen K, Weidler M;

DR WPI; 2002-269010/31.  
xy

PT Identifying plant target proteins for herbicidally active compounds  
PT comprising aligning and comparing nucleic acid or amino acid sequences  
PT from plant with nucleic acid or amino acid sequences from non-plant  
PT organisms -

Claim 5; SEQ ID NO 1505; 261pp + Sequence Listing; English

The invention relates to identifying target proteins (AB990790-AB994016) for herbicidally active compounds, comprising aligning and comparing nucleic acid or amino acid sequences from plant with nucleic acid or amino acid sequences from non-plant organisms using suitable search parameters, where plant sequences having an E-value greater by a factor of 3 than the E-value of most similar non-plant sequences are selected. The polypeptides or nucleic acids encoding them are useful for identifying modulators. The identified modulators are useful as herbicides.

**SQ**      **Sequence**      **1164 AA;**

Query Match	42.6%;	Score 2633.5;	DB 23;	Length 1164;
Best Local Similarity	48.8%;	Pred. No. 4e-191;		
Matches 568; Conservative	182;	Mismatches 296;	Indels 119;	Gaps 27

[illegible]

```

Db 107 YLGNFS-----SGDSSSSS-----GC----- 124
QY 165 KVGGLKLNLEVTLDLSANISGANVYGVWLSDGCGELKHLAISGNKISGDVDVSRVN- 223
Db 125 -----SLEVLVDLSNLSLTDSSIVDYVST-CLNLVSVNFSHNKLAKGLASSPSAN 174
QY 224 --LEFLDVSSNFSFGIP--FLGDC--SALOHLDISGNKLSGDFSR----- 263
Db 175 KRITTVLSDNNRFBDELPETFIADFPNSLKHLDLSGNVNVGDFSRSLFGLCENLTVFSLS 234
QY 264 -----AISTCELKLNLSNQFVGPPIPLP-----LKSQVLSLAENKFGCEIP 308
Db 235 QNSISGRFPVSLNCKLLETNLNRSNLIGKIPGDDYGNFQNLRLSLAHNLVYSGEIP 294
QY 309 DFLGACDCTLGLDLSGNHFGAVPFGSCSLLESLSLSSNNSGGLPMDTLKMRGLK 368
Db 295 PELSLCETLEVLDSGSLTGQLPQSTSCGSLQSLNLKLNKLSGDFLSTVYSKLSRIT 354
QY 369 VLDSFNEFSGELPELSTNLASLTLTLDLSNNSGPIPLNLCQ-NPKNTLOELYLQNG 427
Db 355 NLVLPFNNSISVPSLSTNGS--NLRVLDLSNFBTEGVPSGFCSLQSSVLEKLLIANNY 413
QY 428 FTGKIPTLNSCELSVLSLSPNLSGTPISLSGLSLKLDLKLMLLSEIPEQL-MY 486
Db 414 LSGTVPVELGCKSLKTLTDLSPNALTGLIPKEIWTLPGLSDLVMMANNLTGIPESICVD 473
QY 487 VKTLETLIDFNDLTGEIPSGLSNCTNLNWLISLNNRLTGEIPKMGRLNLAIKLSNN 546
Db 474 GGNLETLILNNTLGLSLPEISLCKTNMTLISLNNLTGELPVGIGLEKALILQNGN 533
QY 547 SPSGNIPELGDCLSLMLDLNTLNFSTIPAMFKGSGKLAANFIAGKRVYIKNDGK 606
Db 534 SLTGNIPSELGCKNMLIWLDLNNSNLTGMLPEGLASQAGLVWPGSVSGKOPAFVANE-G 592
QY 607 KECHGAGNLEFQGRSDQANLSTRNCONITSRVYGHSTPFPNNSMFLPMYSYML 666
Db 593 TDCRAGGIVVEGRFAERLEHFPVHSCPKT-RIYSKMTMTMESNSGSMIYLDLSTAY 651
QY 667 SGYIPEKISMYPLFLNLGNHDSIGSIPDEVDLRGNLIDLSNKLDRIPQMSALT 726
Db 652 SSSIPGLYAMGYLQVNLGNHLLTGTIPDSFGKALGVLDLSHNDLQGFPLSGLSLS 711
QY 727 MUTEIDLNNNLISGPIPEMGOFETPPPAFLNNPGLCGYPLRCPSNADGVAHHQSHG 786
Db 712 FLSDLDVSNNNLTGPPIPGGOLTPPLTRYANNISGLCGVPLPCCSG-----SRPTRSHA 766
QY 787 R-RPASLASVMMGLLFSFVCI FGLTVGRNKRKRKEALEMYAEHGNSGDRTAN 845
Db 767 HPKQSIATGMSAGIVSFMCIVMLIMALYAPAK-VQKEXQREKXIIBSLPTS-----S 820
QY 846 TNNKLTGVKALISINLAIEKPLKLTPLADLQATNGFINDSLISGGGVDYKALIDG 905
Db 821 SSMKLSVHEPISINATEPEKPLKLTFAHLLEATNGFSADSMISGGGPDYKAKLADG 880
QY 906 SAVAIAKLIVHSGQGRPEPAEMETIGKIKHRLVPLGLCKIGERLLVYEMKGSLE 965
Db 881 SVAIAKLTQVTOGQGREFAEMETIGKIKHRLVPLGLCKIGERLLVYEMKGSLE 940
QY 966 DVLOD--PKKGVKLTSTRKRIAGSARGLAFHNCSPIIHRDMKSSNVLLDENTLEAR 1024
Db 941 TVLHEKTKKGGIFLDMARKKIAIGAARGLAFLHSGIPHIHRDMKSSNVLLDQFVAR 1000
QY 1025 VSDFGMARLMSAMDTLISVSTLAGTGYVPPEYQSGRCSTKDYVSYGVVLELTLTGR 1084
Db 1001 VSDFGMARLVSALDTLHLSVSTLAGTGYVPPEYQSGRCSTKADYVSYGVVLELTLTGR 1060
QY 1085 PTDSPDG--NNNLVGVWKO--HAKLRISDVDPPELMKEDPALTEIELQHLKVAVACLDDBA 1142
Db 1061 PTDPEEGEDNNLVGMKOLYREKGAHIDPELV--TDKSGDVELHLYKLASQCLDDRP 1119
QY 1143 WRRETVQVWAMPEK--IOAGSGIDS 1166

```

```

Db 1120 FKRPMTIQVMTMFKELVQVDTENDS 1144
RESULT 6
ABB91366
ID ABB91366 standard; Protein; 1166 AA.
XX
AC ABB91366;
XX
DT 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 577.
XX
KW Herbicidal; plant; agriculture; herbicide.
XX
OS Arabidopsis thaliana.
XX
PN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP09892.
XX
PR 28-AUG-2001; 2001WO-EP09892.
XX
PA (FARB ) BAYER AG.
XX
PI Tietjen K, Weidner M;
XX
DR WPI; 2002-269010/31.
XX
PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX
PS Claim 5; SEQ ID NO 577; 261pp + Sequence listing; English.
XX
CC The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX
SQ Sequence 1166 AA;
XX
Query Match 42.4%; Score 2623.5; DB 23; Length 1166;
Best local similarity 48.5%; Pred. No. 2.3e-190;
Matches 573; Conservative 190; Mismatches 341; Indels 77; Gaps 29;
QY 33 YRETHOLISRK--DYLPD--KNLLPDW--SSNKVPCFEDVTCDD--KVTSIDLSKPLNV 86
Db 32 FNEYTALLAKRQNSVKSDEPNNVLGNNMKYRSGSCSMRGVSCDDGRIVGLDRLNGLTG 91
QY 87 GFSVSSSLSLTGLTESLFLSNSHIN--GSVSGFKCSASLTSIDLSRNSLSGFTTLTSL 144
Db 92 TLNLV--NLRLPQLQVLYIQGNVFSGSDSDC--YQVLDLSNLSIDYSMDYVF 147
QY 145 GSCGLKFLVNSNTLDFPKVVS--GGLKNSLSEVLDSLANSISGANVYGVWLSDGCGELK 203
Db 148 SKCSNLSVSNISNNKL--VGKLGAPESLQSLTTLVDSYNTLSD--KIPESFISDFPASKL 204
QY 204 HLAISGNKISGDVAVSVCNMLEFLDVSSNNFSTGIPFLGDCSALOHLDISGNKLSGD--FS 262
Db 205 YLDLTHNNLSGD-----FSDLSS-----FGICNLTFFSLSQNNLSGDKFP 244
QY 263 RAISCTEELKLNLSNQFVGPPIPLP-----LPLKSQVLSLAENKFGCEIPDFLSGACDTL 318
Db 245 ITLPNCKLELTNLSRNNLAKGIPNGEYWGSPQVLRKQSLAHNRLSGEIPPELSLCKTL 304

```

QY 319 TGLDLSGNHFGVAVPFFGSCSLLESLALSSNNNSGELPMDTLTKMRGLKVLDSFNEFS 378  
 DB 305 VILDSGNTFSGELPSQPTACVWLQNNLGNVYLSGDELNTVSKITGITVLYAVANNIS 364  
 QY 379 GELPESLTNLASLTLLDLSNNFSGPILPNILO-NPKNTLOEYLQNGFTGKIPPLS 437  
 DB 365 GSVPLSTLNGS-NLRVLDLSSNGFTGVPSPGCSLQSSPULKEKILIANVYLSGTVPELG 423  
 QY 438 NCSELVSLHSPNYLSGTPSSLSGLSKLDLKLMLNMLEGEIPELMTYK-TLETLIL 495  
 DB 424 KCKSLKTIDLSFNNELTGPDKPEIWMPLSDLVMMANNLTGTIE-EGCVYCGNLETLIL 482  
 QY 496 DFNLTGEPISGLSNCCTMLNWSLSNNRLTGEIPKATGLENLAILKLSNNSFGNIPDE 555  
 DB 483 NNNLTGSPISIRCTNMWISLSNNRLTGPISGIGLSKALITLOLGNNSLSGNVPRQ 542  
 QY 556 LGDRSLIWLDTNLNFGTIPAMFQSGKIANFLAGKRYVYIKNDGKKECHGAGNL 615  
 DB 543 LGNCKSLIWLDTNLNFGTIPAMFQSGKIANFLAGKRYVYIKNDGKKECHGAGNL 601  
 QY 616 LEFGIASEQNLSTNPNCTISRYVYIGHTSPTEGNNSSMFLDMSYNNLSGYIPEIG 675  
 DB 602 VEFEGIAERLERLPMVHSCPAT-RIYSGMTMYFSANGSMIYFDISYNAVGPPIPGYG 660  
 QY 676 SMPYLFITINIGNDISGSIPEVGDRLGMLNIDLSNKLDRIPQMSALTMTEIDLSN 735  
 DB 661 NMGLQVLTNIGNHRITITIDSPFGSLAIGVLDLSHNNLQGYPSGLSLSFISLSDIVSN 720  
 QY 736 NNTSGPIPEMGQETFPFAKFLNPNGLCGVPLPCDPSNADGVAHHQSRGRPASLAGS 795  
 DB 721 NNTLGPPIFGQULTFPVSRYANNISGLCGVPLPCDPSNADGVAHHQSRGRPASLAGS 777  
 QY 796 VANGLSFPCITGILLVGRMKRRKKEAELEMAEAGNSGDRNTANTNKLTVKYE 855  
 DB 778 VIAGIASFPCFVWLVALYVRK-VQKKEQKREKYEISLPTSG-----SCSWKLSVPE 831  
 QY 856 ALSINLAPEKPKLTFADLQATNGFNHDSLISGSGFDVYKALIKGSAVAIKLH 915  
 DB 832 PLGINVATPEKPKLTFADLQATNGFNHDSLISGSGFDVYKALIKGSAVAIKLH 891  
 QY 916 VSGQGRBFAENETTGKIKHNLVPLLYGCKVYDERLLVNEVMKYSLEVDLQD--PKK 973  
 DB 892 ITQCGRBFNAENETTGKIKHNLVPLLYGCKVYDERLLVNEVMKYSLEVDLQD--PKK 951  
 QY 974 GGVKLTSTBRKAISANGELFLHNCSPHITHRMKSNVLLDENLFAVYSDFGMARL 1033  
 DB 952 GGVKLTSTBRKAISANGELFLHNCSPHITHRMKSNVLLDENLFAVYSDFGMARL 1011  
 QY 1034 MSAMDTLHSTLACTGPGVVPPEYQSFRCSTKGVYSGVYVLELITGKRPDSDPFG- 1092  
 DB 1012 VSMALDTLHSTLACTGPGVVPPEYQSFRCSTKGVYSGVYVLELITGKRPDSDPFG- 1071  
 QY 1093 DNNLVGVWYQ-HAKLISVFPBELMKEDPALEIELLQHLKVAVALCLDDRAMRPTMYOV 1151  
 DB 1072 DNNLVGVWYQ-HAKLISVFPBELMKEDPALEIELLQHLKVAVALCLDDRAMRPTMYOV 1130  
 QY 1152 MAMFEKIQAGSGIDSGSTIRSIJEDGFSFTEMDMSIKKVP 1192  
 DB 1131 MAMFEKEMKADTEDE-----SLDEFSLKXTP 1156

RESULT 7  
 ABB91692  
 ID ABB91692 standard; Protein; 1143 AA.

AC ABB91692;  
 XX 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 903.  
 XX Herbicidal; plant; agriculture; herbicide.

XX Arabidopsis thaliana.  
 OS  
 XX  
 PN MO200210210-A2.  
 XX  
 PD 07-FEB-2002.  
 XX  
 PF 28-AUG-2001; 2001WO-EP09892.  
 XX  
 PR 28-AUG-2001; 2001WO-EP09892.  
 XX  
 PA (FARB) BAYER AG.  
 PI Tiejfen K, Weidner M,  
 XX  
 DR WPI, 2002-269010/31.  
 PT Identifying plant target proteins for herbicidally active compounds,  
 PT comprising aligning and comparing nucleic acid or amino acid sequences  
 PT from plant with nucleic acid or amino acid sequences from non-plant  
 PT organisms -  
 PS Claim 5; SEQ ID NO 903; 261pp + Sequence listing; English.  
 XX  
 CC The invention relates to identifying target proteins  
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising  
 CC aligning and comparing nucleic acid or amino acid sequences from plant  
 CC with nucleic acid or amino acid sequences from non-plant organisms using  
 CC suitable search parameters, where plant sequences having an E-value  
 CC greater by a factor of 3 than the E-value of most similar non-plant  
 CC sequences are selected. The polypeptides or nucleic acids encoding them  
 CC are useful for identifying modulators. The identified modulators are  
 CC useful as herbicides.  
 XX  
 SQ Sequence 1143 AA;

Query Match 40.6%; Score 2509.5; DB 23; Length 1143;  
 Best Local Similarity 46.2%; Pred. No. 1.1e-181;  
 Matches 556; Conservative 183; Mismatches 353; Indels 111; Gaps 25;

QY 9 LSVTLTFFSFFSLSPQASPSQSIYREIHQISFKVLPD-KNLEPDMSSNKPCTFDG 66  
 DB 14 IQISIFILTLHLSQS-SSDQSSILKTSLSLSPKTIQDDPNNTLSNWSPPKQFSG 72  
 QY 67 VTCRDVYTSIDSSKPLN-VGSAVS-----SLSLTGLSFLSNHSINSVSGFKC 120  
 DB 73 VTCRGVYTNLSGSLSGIVSFNATSLDSLVKLS-ENFVYLN-----STSLILL 125  
 QY 121 SASLTSIDLSRNSLSGVTTLTSLGSCSGKFLNVSNTL--DFPKYSGKULNS--L 175  
 DB 126 PLTTLHLESSGLIG-----TLPENFFSKYGNLISITLVSNNFTGKLPNDLFLSSKUL 179  
 QY 176 EVDLSANSISGANVGVWVLSGCGELKHLAISGNKISGVYVSRVNLFEFLDVSSNNS 235  
 DB 180 QTLDSLNNITG-----PISGLTIP----- 199  
 QY 236 TGLPFLDCSALQHLDSGNKLSGDSFRAISTCTEKLINISNOVGPPIPL-PLKSL 293  
 DB 200 -----LSSCVSMYTYLDRSGNSISGYISDLSINCTNLSKSLNNSNDGQIPKSGKLKL 254  
 QY 294 QYLSIENKFTGEIPDFLSGACDTLTGLDSGNHFGVAVPFFGSCSLLESLALSSNNS 353  
 DB 255 QSLDLSNNRLTGMIPPEIGTCTSLQNLRLSYNNFTGVPISLSSCSMWQSLDLSNNSIS 314  
 QY 354 GELPMDTLLMRGLKVLDSFNEFSGELPESLTJNLSA--SLTLTDLSSNNSFGPILPNC 411  
 DB 315 GPPPNLTILRSFSLQIULLSNNLISGDFP---TSISACSLMIADSSNRFSGVLPDLC 371  
 QY 412 QNKNTLOEYLQNGFTGKIPPLTSSNGSELVSLHSPNYLSGTPSSLSGLSKLDLKL 471  
 DB 372 PGAA-SLEBLRFPDNLVTEIPPAISQCELRITIDLSLNYLNGTIPPEIGNQKLEQFIA 430  
 QY 472 WLNMLEGEIPQELMYVKTLETILDPNDLTGELIPGSLSNCCTMLNWSLSNNRLTGEIPRW 531







QY 481 POELMYYKTLFTLLDPRNDLTGEIPSGISNCTNLMNISNNRLTGEIPKIGLEMLAI 540  
 Db 434 PPELJCKCNLKNLLNNHNLKGEIPTELFCNSMLEWISLTSELTGEVREVLRLAV 493  
 QY 541 LKLSNNSPSGNIPELGDRCRLIWLDTNLTNLFNGTIPAAAFKSG-KIANPFLAGRYV 599  
 Db 494 LQLNNSISGQIPPELNCSSIMLWDLSSNKLGEIPRLGRQAKGPGIPSGTIVF 553  
 QY 600 IKNDGKKECHGAGNLLFQGISBQNLRLSTNPNCTISRVVGHSTSPFDNNGSMFL 659  
 Db 554 VRNVG--NTCKGVGGLLEFAGIRPERLLQVPSLRTCNF-ARMYSGPILSKFTTYQVEVL 610  
 QY 660 DMSYNMLSGYIPKESGMPYLTINLGHNDISGSPDEYVDLFGNLTDLSSNKIDGRIP 719  
 Db 611 DLSNQRKGIPEFEGDVALQVLEISYNQSGIPSSIGRLKDLGVFAASHRLQGGIP 670  
 QY 720 QAMSAITMLTEIDLSNNNLSGPIPEMGOFTFPFAFLNPPGCGVPLRRC-----DP 772  
 Db 671 DSFSNLSFLVQIDLSYNDLTGQIPQGGQSTLPASQYENHPGLCGVPLPECRNNEPETNP 730  
 QY 773 SNADGVANHQRSGR-RP-ASLAGSVANGLLFSFVCIFGLILVGRNMRKRKKEAELE 829  
 Db 731 DVAMG-----TGRAKPGTASWANSIVIGIILISASICILTWGIAKRSRKEAF-EMK 783  
 QY 830 MYAEGHNSGDRFANNNTNMTLGVKALSTINLAPEKPIKLTPLADLQATYGFNDSLI 889  
 Db 784 ML-----KSLQAAHATTWKIDKEKEPLSINAVTPQRRLKLFQSLLEATNGFSABSLI 838  
 QY 890 GSGGFGDVAIKKDSNAVAIKKLHVSGQGRFEMAEWETIGIKHNILVPLGYCYVG 949  
 Db 839 GCGGFSEVFATKDKGSNAVAIKKLIRLSTVQGRFEMAEWETIGIKHNILVPLGYCYVG 898  
 QY 950 DEELLVNEWKVSGLEEDVQDPKGGV-----LKLSTRKIAIGARGLAFLHNCSP 1003  
 Db 899 KEHLVYEFERGESLEEMH---GRVKTQDRRIITWEERKKIARGAAGKCFLEHNCIP 954  
 QY 1004 HIIHRMKSNNVLDENLEFARVSDFGMARLMSANDTHLSVSTLACTPGVPEVYQSRIC 1063  
 Db 955 HIIHRMKSNNVLDHDMARVSDFGMARLISALDTHLSVSTLACTPGVPEVYQSRIC 1014  
 QY 1064 STYGDVYSIGVVLLELTGKRPDTPSPFGDNNLVGVYQHAKL-RISDVPPELM----- 1117  
 Db 1015 TAKGDVYSIGVVLLELTGKRPDTPSPFGDNNLVGVYQHAKL-RISDVPPELM----- 1117  
 QY 1118 KEDPALEI-----ELLQHLKVAVALCLDRAKRRPTMVAAMFEIQAQS-GIDSGST 1169  
 Db 1075 QEBARAEAAEAVVKEWRYMETITLQCVDPFSRRPTMLQVYAMRELVPGSTVSSNST 1133  
 KDEL 9  
 AAB25323  
 ID AAB25323 standard; Protein; 919 AA.  
 AC AAB25323;  
 XX  
 XX 27-NOV-2000 (first entry)  
 DE Bucalypus grandis cell signalling involved protein SEQ ID NO:642.  
 XX  
 XX  
 KM Bucalypus grandis; Pinus radiata; Monterey pine; plant; modification;  
 KM plant cell signalling; modulation; transgenic plant; pathogen; growth;  
 KM environmental change; development; cell proliferation; differentiation;  
 KM elongation; survival; disease resistance; nutrient metabolism.  
 OS  
 XX Bucalypus grandis.  
 XX  
 PN MO200042171-A1.  
 XX  
 XX 20-JUL-2000.  
 PD 11-JAN-2000; 2000MO-US00724.  
 PF  
 XX 12-JAN-1999; 99US-0228986.  
 PR

PR 01-NOV-1999; 99US-0162866.  
 XX  
 PA (GENE-) GENESIS RES & DEV CORP LTD.  
 XX  
 PI Strabala TV, Nieuwenhuizen NT;  
 XX  
 XX WPI; 2000-476052/41.  
 PT Isolated polynucleotide encoding a polypeptide involved in cell  
 PT signaling used for generating transgenic plants with modified responses  
 PT to external signals -  
 XX  
 XX  
 PS Claim 3; Page 291-293; 527pp; English.  
 CC AAA79263 to AAA79736 and AAB25100 to AAB25570 represent polynucleotide  
 CC and protein sequences isolated from eucalyptus (Eucalyptus grandis) or  
 CC pine (Pinus radiata also known as Monterey pine). The protein sequences  
 CC are involved in cell signalling. The polynucleotide and protein  
 CC sequences can be used to modify the response of plant cells to external  
 CC signals e.g. environmental changes or pathogens during the growth and  
 CC development of a plant. They can be used to modify cell proliferation,  
 CC differentiation, elongation and survival, resistance to disease and  
 CC nutrient metabolism. Examples of modifications which can be produced are  
 CC altered fruit ripening and senescence of leaves and flowers e.g. to  
 CC delay senescence and prolong the life of cut flowers or enhance  
 CC modifications of reproductive organs to engineer sterile plants. Other  
 CC organs providing fruit and vegetables which have a longer shelf life  
 CC between harvest and consumption, or to decrease branching frequency in  
 CC forest tree species giving long stretches of valuable knot-free clear  
 CC wood which can be used in solid timber furniture and veneers.  
 XX  
 XX Sequence 919 AA;  
 SQ  
 Query Match 35.4%; Score 2189.5; DB 21; Length 919;  
 Best Local Similarity 50.6%; Pred. No. 1.7e-157;  
 Matches 475; Conservative 134; Mismatches 276; Indels 53; Gaps 18;  
 QY 264 AISTCETELKLNINSGVGPPIPL--PLKSLQYLSLAENKFTGEIPDLGACDTLTGL 321  
 Db 3 SFSNCTSLHTLSLANNFTGEIPHSFGRSLASLQKLDLSHNLHGMVDELGTACNSLIEL 62  
 QY 322 DLSGHHFYGAVPPFGSGSLLESIALSSNNFSGELPMDTLTKRGLKVLDTSPNERSGEL 381  
 Db 63 TLSFNNISGPIPLSPSSGSMWQSIDLSTNNISGEPFDSILRNLGSLSLSNNTTSGLF 122  
 QY 382 PESLTNLSASLTLDLSSNFGPIPLNLCQNPKNLTQELYLQNGFTGKIPTLSCNSB 441  
 Db 123 PASL-SYCKKRLTIIDFSSNCFSGIIPDLCLGAA-LLEIRAPADNLTIGNIIPQLSQCSQ 180  
 QY 442 LVSLHLSFNYLSGTIPSSLSGLRDLKULMLNLEGEIPQELMYKTLFTLLDPRNDLT 501  
 Db 181 LKIDPFLNLYVGSIPAEIGLNTLEQLIMFNGLGEIPPELGKRCNKLNNHNLK 240  
 QY 502 GEIPSGLSNCTNLMWISLNNRLTGEIPKWIIGLEMLAIKLSNNSPSGNIPELGDRCRS 561  
 Db 241 GEIPTELFNCSNLEWISLTSELTGEVREVLRLAVLQGNNSISGQIPPELNCSSIML 300  
 QY 562 LIMPDLMTNLFNGTIPAAAFKSG-KIANPFLAGRYVYIKNDGKKECHGAGNLLFQGIS 620  
 Db 301 LMDLSSNKLGTGIPRLGRQAKGPGIPSGTIVFVRNVG--NTCKGVGGLLEFAG 358  
 QY 621 IRSBQNLRLSTNPNCTISRVVGHSTSPFDNNGSMFLDMSYNMLSGYIPKESGMPYLT 680  
 Db 359 IRPRLQVPSLRTCNF-ARMYSGPILSKFTTYQVEYLDSDNQLRGKIPEFEGDVAL 417  
 QY 681 FIILNLDNDISGSPDEYVDLFGNLTDLSSNKIDGRIPQAMSAITMLTEIDLSNNNL 740  
 Db 418 QVLESLYNQSGIPPELNCSSIMLWDLSSNKLGEIPRLGRQAKGPGIPSGTIVF 477  
 QY 741 PIPEMGOFTFPFAFLNPPGCGVPLRRC-----DPSNADGVANHQRSGR-RP-AA 790  
 Db 478 QIPQGGQSTLPASQYENHPGLCGVPLPECRNNEPETNPDVAMG-----TGRAKPGT 531

QY 791 SLAGVAMGILFPGVCIPLGLIIVGREMRKRRKKEALEMYAEGHNSGSDRTANNIMKL 850  
 Db 532 SWANSIVLGLISIASICILTVGIMARSRRKAE-EMKML-----KSIQAAHATTWKI 585  
 QY 851 TGVEALISIMAAFEKPLKLTPEADLLQATNGPHNDLSIGSGFDVYKAILKDSAVAI 910  
 Db 586 DKKEPELSIVAFQGRRLKRLKFSQILEATNGFSAESLIGCGFGVFAATLKDGSNVAI 645  
 QY 911 KXLIHVSQGDREPMAMETIGIKIRNLVPLIGYCKVGBERLLVMEVMKYSLEPVLQD 970  
 Db 646 KXLIHVSQGDREPMAMETIGIKIRNLVPLIGYCKVGBERLLVMEVMKYSLEPVLQD 704  
 QY 971 PKGKGVK-----LKTSTRKIAIGSARGLAFLHNSCPHIIHRDKSSNVLLDENLEAR 1024  
 Db 705 ---GRVITQDRRLITWEERKKIARGAKGICPLHNCIPHIHRDKSSNVLLDHMEAR 761  
 QY 1025 VSDGFMARLMSAMDTLVSSTLACTPGVPEYVQSFRCTKGDVYSGVVLLELTGKR 1084  
 Db 762 VSDGFMARLISALDTHLSVSTLACTPGVPEYVQSFRCTKGDVYSGVVLLELTGKR 821  
 QY 1085 PTSDPDGDNVLVGMVQMAHL-RISDVPEBELM-----KEDPALER-----ELLOHLK 1132  
 Db 822 PTDEDREDTDLVGMVQMAHREGKQMEYIDPAMLSKTKVGBEALAEAVVEMRYME 881  
 QY 1133 VAVACLDDBRAMRPTVMQVMAFKEIQAGS-GIDSQST 1169  
 Db 882 ITLQCVDDFPSPRRPTMLQVAMLELIVPGSTTVSSNST 919

RESULT 10  
 ABB93388  
 ID ABB93388 standard; Protein; 1192 AA.

AC ABB93388;  
 XX 31-MAY-2002 (first entry)  
 XX 31-MAY-2002 (first entry)  
 DE Herbicidally active polypeptide SEQ ID NO 2599.  
 XX Herbicidally active polypeptide SEQ ID NO 2599.  
 KW Herbicidal; plant; agriculture; herbicide.  
 OS Arabidopsis thaliana.  
 XX Arabidopsis thaliana.  
 XX WO200210210-A2.  
 PD 07-FEB-2002.  
 XX 28-AUG-2001; 2001WO-EP09892.  
 XX 28-AUG-2001; 2001WO-EP09892.  
 PR 28-AUG-2001; 2001WO-EP09892.  
 XX (FARB ) BAYER AG.  
 XX (FARB ) BAYER AG.  
 PA Tietjen K, Weidner M;  
 XX Tietjen K, Weidner M;  
 DR WPI; 2002-269010/31.  
 XX Identifying plant target proteins for herbicidally active compounds,  
 PT comprising aligning and comparing nucleic acid or amino acid sequences  
 PT from plant with nucleic acid or amino acid sequences from non-plant  
 PT organisms -  
 XX Claim 5; SEQ ID NO 2599, 261pp + Sequence listing; English.  
 XX The invention relates to identifying target proteins  
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising  
 CC aligning and comparing nucleic acid or amino acid sequences from plant  
 CC with nucleic acid or amino acid sequences from non-plant organisms using  
 CC suitable search parameters, where plant sequences having an E-value  
 CC greater by a factor of 3 than the E-value of most similar non-plant  
 CC sequences are selected. The polypeptides or nucleic acids encoding them  
 CC are useful for identifying modulators. The identified modulators are

CC useful as herbicides.  
 XX Sequence 1192 AA;  
 SQ  
 Query Match 26.0%; Score 1608.5; DB 23; Length 1192;  
 Best Local Similarity 35.2%; Pred. No. 4.4e-113;  
 Matches 453; Conservative 177; Mismatches 421; Indels 235; Gaps 36;  
 QY 11 VTTLFFSFFSLSQASPSQSLREIHQSLSPFDVLPDKLLLDW--SSNKPCTPDGVT 68  
 Db 4 LTALFLFLFF--SPSSSAIVLSETTSLISPKSLNPSLSMNVSSASACDVGVT 61  
 QY 69 CRDKVMSIDSSKPLWVGSFAVSSLSLTGLESLPLSNHNGSVS---GFKCSASL 124  
 Db 62 CLIRGVNSLSPSLSLR---CQIPKEISLKNLRELCIAGNQSFKIIPPEIMWLK--HL 115  
 QY 125 TSLDSRNSLSGPVTTTSLSGSGGLKFLVNSVNTLDPFKV--SGGLKLSNLEVLISA 182  
 Db 116 QTLDSGNSLTGLLPRL--LSELPLQLYLDLSDN--HFGSLPPSPFTSLPALSSLDVSN 171  
 QY 183 NSISGANVGVLDGCGELKHLA--ISGNKISGV----- 216  
 Db 172 NSLSGE-----IPPEIGKLSNLSNLVGLNSFGQIPSEIGNISLKNFAPECFENG 225  
 QY 217 ---DVSRCVNLFFLDVSSNNPSTGI-----PFLGDCSALQ 248  
 Db 226 LPKEISKXHLAKLDLSYNPCKSIPKSGFGLHNLITLIVSAELIGLIPBELCNCKSLK 285  
 QY 249 HUDISGNKLSGDFSPRAISTCHELKILNLS--NQFVGPIPL--PLKSLOYLSAENKFT 304  
 Db 286 SIMLSFNSLSGPLPLELS--EIPLLTFSAERNQSGSLPMSMGKVKYLDLILANNRFS 342  
 QY 305 GRIP-----DPLSGA-----CDTLTGDLDSGNHFFYGAVPFFGSCSL 341  
 Db 343 GELPHEIEDCPMLKXLSLANSLSGSIPELQSGSLAIDLSGULSGTIEEVEDGSS 402  
 QY 342 LESLASSNNFSGELPMDTLKMRGLKYLDLSFNEFSELPESL--TNL----- 388  
 Db 403 IGEILLITNNQINGSIPED-LMKL-PLMALDLSNNFTGEIPSKSTNLWEFTASYNRL 460  
 QY 389 -----SASILLTDLSSNNFSGPILPNLCQPKXTLOELYLONNGFTGKIPTLSN 438  
 Db 461 EGYLPKEIGNAASLKLVLVSDNQLTGEIPEI--GKLSLSTVLNANAMPQKXIPVELGD 518  
 QY 439 CSELVSLHSFNYLSGTPISLSGLSKLDLKLWNLGEBIPQ-----ELMY 486  
 Db 519 CTSLTTLTDLGSNNLQGIQDXTALAQCLVSYNNLSGSIKSPSAFYHOIEMPDLSF 578  
 QY 487 VKTLETLIIDFNDLTGEIISGLSNCNTNLMWISLNNRLTGEIPEKXIGRENLATILKSN 546  
 Db 579 IOHHGIFDLSYNNLSGPIPEELGECLVLEISLNNHLSGELPASISRLTNLTIIDLSGN 638  
 QY 547 SPSGNIPDELGDCRSLLIMDLNLTNLFNGTTPAMFQSGKIANPFAKRYVYIKNDGK 606  
 Db 639 ALTGSIPKEMGNSLAKQGLNLANQNLGHI----- 669  
 QY 607 KECHGANLLEFOGIRSEQLNRLSTRNPNITTSRYVGGTSPFEDNGSMFLDMSYML 666  
 Db 670 -----ESFGILGSLVKLNTLNKLDGVPASIGLKLTHNDLSFNNL 712  
 QY 667 SGYIPKESGMPLFILNLGHNISGSIPEVGLDGLNLTDLSSKLDGRIPOAMSAVL 726  
 Db 713 SGEISSELSTMEKLVGLYIRQNKFTGIBSELNLQLEIYDLSNELLGELIPTKICGLP 772  
 QY 727 MLTEIDSSNNNLGGPIPEMQGFETFPAPKLANPGLGCVPLPCDPSNADGAHQRSHG 786  
 Db 773 NLEFLANAKNNLGEVSDVQDPSKALLSGNKEICG-----RVVGSDC-----KING 821  
 QY 787 RRPASLAGVAMGILFS--FVCIPLGL--ILVGREMRK--RRKKEALEMYAEG--HG 836  
 Db 822 TKLASAVGIAGMLGFTIIVFVFSLRRAWMTKRYQQRDPERMEBSRLKGVVDQNLV 881  
 QY 837 NSGRTANNTWKLTGVKALISIMAAFEKPLKLTPEADLLQATNGPHNDLSIGSGFGD 896

```

Db      882  LSGSS-----REPUSINAMEEOPPLKVRIGDIVEATDHPFSKNIIGDGGFGT 930
Qy      897  YKAILKDGSAVAIKLTHVSGQDREFMAEMETIGIKRNLVPLIGYCKVDEBLTN 956
Db      931  YKALCPBEKTVAVKLSEAKTQGNREFMAEMETIGYKRNVLVSLGYSFSEKLLVY 990
Qy      957  EVMYKGSLEDVLOPKKGVKALKSTRKLAIGSARGLAFIHNCSPHIIRDMKSNVY 1016
Db      991  EYMNNGSIDHMLKQTMLEVLDMWSKRLKIAVGARGLAFIHHGFIPIIHRDIKASNTL 1050
Qy      1017  LDENLEAVSDFGMARLMSAMDTHLSVSTAGTPGYVPEPYGQSFCSKTDVYSGVYL 1076
Db      1051  LDGDFEPKVNADFGARLISACEHSV--TVIAGTGTGIIPEPYQSARATYKGDVYSFGVIL 1109
Qy      1077  LELITGRPTDSDPFGGN--NLVGVWKQHAU--RISDVEPELMKEDPATELELQHLK 1132
Db      1110  LEIYTGKEPT-GPDFKESEGNLVGMAIQKINGQKAVDVIDPLV--SVALKNSQLRLIQ 1166
Qy      1133  VAVACLDDBAMRPTMOWAMPEKEI 1158
Db      1167  IAMCLAEPAKRPMDVLKALKEI 1192

```

RESULT 11  
 ABB93804  
 ID ABB93804 standard; protein; 1036 AA.

AC ABB93804;  
 DT 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 3015.

KM Herbicidal; plant; agriculture; herbicide.

OS Arabidopsis thaliana.

PN W0200210210-A2.

PD 07-FEB-2002.

PF 28-AUG-2001; 2001WO-BP09892.

PR 28-AUG-2001; 2001WO-BP09892.

PA (FARB) BAYER AG.

XX Tietjen K, Weidner M;

DR WPI; 2002-269010/31.

PT Identifying plant target proteins for herbicidally active compounds,  
 comprising aligning and comparing nucleic acid or amino acid sequences  
 from plant with nucleic acid or amino acid sequences from non-plant  
 organisms -

PS Claim 5; SEQ ID NO 3015; 261pp + Sequence Listing; English.

CC The invention relates to identifying target proteins  
 (ABB90790-ABB94016) for herbicidally active compounds, comprising  
 aligning and comparing nucleic acid or amino acid sequences from plant  
 with nucleic acid or amino acid sequences from non-plant organisms using  
 suitable search parameters, where plant sequences having an E-value  
 greater by a factor of 3 than the E-value of most similar non-plant  
 sequences are selected. The polypeptides or nucleic acids encoding them  
 are useful for identifying modulators. The identified modulators are  
 useful as herbicides.

XX Sequence 1036 AA;

Query Match 22.7%; Score 1402.5; DB 23; Length 1036;  
 Best Local Similarity 33.4%; Pred. No. 1.7e-97;

```

Matches 394; Conservative 182; Mismatches 416; Indels 189; Gaps 33;
Qy      11  VTLFFPSFSLFPQASPSQSLYREIHOLISFDVLPDQULLPDMSSNNPCTPGVTCR 70
Db      4   ILLVFFVSGSVSQPCHPN-----DISALRELAGALKKNSVRESMLNRCCEMVGJCE 58
Qy      71  ----DDKVTSIDLSSKPLVNGFSAVSSSLITGLESLFLSNHSHNGSVSGKCSASLTS 126
Db      59  GSDVSGVTVTLVPEKGL--GVISKSLGELT-----ELRV 92
Qy      127  LDISRNSLSGPTVTLTSLGSCSLKFLANVSNTLPFGKVSGLKLNLEVLDSANSTLS 186
Db      93  LDISRNLKGEV-----PAEIS--KLBOQLVLDLSHNLIS 125
Qy      187  GANNVGVNLSDCGGLKHLAISGKISGV--DVSRCVNLEFLDVSNNPSTGI--PFLGDC 244
Db      126  G-SVLGV--SGKLQISLNTISNSISGKLSDVGFGLVMNVANNLFEGBIHPEL--C 180
Qy      245  SA---LQHLDISGNKLSGDFSRASISCTELKLNLSNSQVFGTIP--LPLKSLQYSLA 299
Db      181  SSSGGIQLVDLSNRLVGNLDGLYNCSKSIQQLHDSNRLTGQLPDYYSIRELOSLIS 240
Qy      300  ENKFTGEIPDFLSGACDTLTGLD--LSGNHFGYAVPPFGSCSLLESIALSNPFGEL 356
Db      241  GNYLSGELSKNLS---NLGKLKSLISENRSDVTPVFGNLTQLEHLDVSNKPFSGRF 296
Qy      357  PMDTLKMGKLVLDLSFNEPFGELPESLTLNLSASLTLDISNNPFGIIPNLCONPKN 416
Db      297  P-PSLSQCSKLVLDLRNNSLSGSLTNLNFGT--DLCVLDLSNHFSGPLPDSIGHCPK 353
Qy      417  TLQELVQNNGFTGKIPTLSNCSLVSILHFNVLSGTIPSSLSGLSKLRDLKMLNML 476
Db      354  -MKLISLANFRGKIPTLPKQLQSLFLSLSN-----SFVDSSETNVL 398
Qy      477  EGBIPELMTVYTLTLIDRDLTGEIPSGLSNCTNMWISLNNLTGEIPKMGRL 536
Db      399  Q-----HCRNLSTLLSKNFGIEEIPNNVTGFDNLATIALGNCGLRGQIPSWLNCK 450
Qy      537  NLAILKLSNNSPSGNIPELDCRSLLWLDLNTLTFNGTTPAAMFKSGKIANFLAKR 596
Db      451  KLEVLDSNHFYGTIPMTGRMESLFYIDPSNTLTGALPVALTELKNIIRLNGTASQM 510
Qy      597  Y-----YIKNDGKKECHAGNMLEFQGISBOLNRLSTRPCNTSR---VVGHT 646
Db      511  TDSGSIPLYKR-----NKSNGLPYNOVSFPFSIT--- 542
Qy      647  SPTFPNNGSMFELDMSTYMLSGYIPKATGSMPIYFIINLGNDISGIPDEVDLRLGNI 706
Db      543  -----LNNRRLNGTILPEIGRLKEIHMLDLSRNNFTGTIPDSIGLDNLEV 588
Qy      707  LDLSNKLDRIPQAMSLTMTLEIDSSNNLSGPIPMGQFETFPAPKIANPGLCGTP 766
Db      589  LDLSNKLDRIPQAMSLTMTLEIDSSNNLSGPIPMGQFETFPAPKIANPGLCGTP 766
Qy      767  LPRCD-----PSNADGYAHQSHRPPASLAGSVAMGLFSEVTCIFGLILVGERKKR 821
Db      649  DSPCDVLMNNLNFKSGSRNNNGKFGKRSIYVLTSLAIGTILLVYL-----RI 702
Qy      822  RKKAELMVAEAGHNGSGDRTANNMTKLTVGEALS--INLAPEK-PLRKTFTADLLOA 879
Db      703  SRKVD-----DRINDVETTSVSKALGPKRIVLFHSCGCKDLSVELLS 750
Qy      880  TNGHNDLSLGGGFGDGVYKAILKDGSAVAIKLTHVSGQDREFMAEMETIGIKRNL 939
Db      751  TNNFSQANIIGCGGFLGVYKANPPDGSKAAVRLSGDCQOMREFOAEVLSRAHKNL 810
Qy      940  VPLIGYKVDDELLVNVWVKYSLDYVQDPKGGVYKALKSTRKLAIGSARGLAFIHH 999
Db      811  VSLGCKHGNDRLLYSFENSGSLDYMLHERVDGMMTLMVDRLIKIQAAGARGLAYLHK 870
Qy      1000  NCSPHIIRDMKSNVLDENLEAVSDFGMARLMSAMDTHLSVSTAGTGYVPEPYQ 1059
Db      871  VCEPNVIHRDVKSNTLDBKPEAHLDARGLARLPDTHVT--TDLVGTIGYLPPEYSQ 929

```

QY 1060 SFRCSSTKGVSYGVVLLHLLTGKPTD-SDFGDNMLVGMWQ-HAKLRISDFPELM 1117  
 DB 930 SLIATCRGVSYGVVLLHLLTGKPTD-SDFGDNMLVGMWQ-HAKLRISDFPELM 988  
 QY 1118 KEDPALEIHLQHLKVAVACLDPRAMRPTMVMAMFEKI 1158  
 DB 989 -RENNVERTVLEMLKACIDHERRRRPLIEBVTWMLDL 1028

## RESULT 12

ABB93939  
 ID ABB93939 standard; Protein; 1102 AA.

AC ABB93939;  
 DT 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 3150.

OS Herbicidal; plant; agriculture; herbicide.

XX Arabidopsis thaliana.

XX WO200210210-A2.

XX 07-FEB-2002.

XX 28-AUG-2001; 2001WO-EP09892.

XX 28-AUG-2001; 2001WO-EP09892.

XX (FARB) BAYER AG.

XX Tietjen K, Weidner M;

XX WPI; 2002-269010/31.

XX Identifying plant target proteins for herbicidally active compounds,  
 PT comprising aligning and comparing nucleic acid or amino acid sequences  
 PT from plant with nucleic acid or amino acid sequences from non-plant  
 PT organisms -

PS Claim 5; SEQ ID NO 3150; 261pp + Sequence Listing; English.

XX The invention relates to identifying target proteins  
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising  
 CC aligning and comparing nucleic acid or amino acid sequences from plant  
 CC with nucleic acid or amino acid sequences from non-plant organisms using  
 CC suitable search parameters, where plant sequences having an E-value  
 CC greater by a factor of 3 than the E-value of most similar non-plant  
 CC sequences are selected. The polypeptides or nucleic acids encoding them  
 CC are useful for identifying modulators. The identified modulators are  
 CC useful as herbicides.

XX Sequence 1102 AA;

XX Query Match 22.3%; Score 1380.5; DB 23; Length 1102;

XX Best Local Similarity 32.0%; Pred. No. 8.8e-96;

XX Matches 387; Conservative 179; Mismatches 470; Indels 173; Gaps 31;

QY 1 MKTFSFPLSVTLFFSPFSLFQSPQSLYREIHOLISFDVLPD-KNLLPDMSSNK 59

DB 5 MMLAVFFSLILLILI-----SETTGALBEGQYLLEIKSKVDAKONLRMNSND 55

QY 60 N-ECTPGVTC---RDKVTSIDLSSKPLNVGFSVSSSLSLTGLBESLFLNSHINGS 114

DB 56 SVFCGRTGWCNSVSDPEVLSLMSMLV-----SGRLSPS 92

QY 115 VSGFKCSASLTSIDLSRSLSGVPTTTLTSLGCSGLKFLVNSNTLDFGKYSGGIKLNS 174

DB 93 IGGI---VHLKQDLISYGLSGKIP--KEIGNCSSLEILKLNNOFDEIPEIG-KLVAS 146

QY 175 LEVLDLSANSISG-----ANVVGWVLSDDGGELKELA---ISGNKIS 213

DB 147 LEMLIYYNNRISGSLPEVEIGNLISLQVTVYSNNISGQIPRISGNLKLTLSPRAQNMIS 206

QY 214 GDV--DVSRCVNLEFLDVSSNNFSTGP-FLGDCSALQHLDISGNLSDGFSRAISTCTE 270

DB 207 GSLPSEIGGESLMLGLAQNQSLGELPKKIGLKKLSQVILWENEFSGFIPREISNCTS 266

QY 271 LKLNISNQGFGVGPPL--PLKSLQYLSAENKFTGEIPDLFGACDVLTLGDLSCNHF 328

DB 267 LETLALYKNQVGPVPELQDLSLEFLYLRNGLNGTIPREI-GMLSAIEIDFSEML 325

QY 329 YGAVPPFGSGSLLESALSSNNFSGELPMDTLKMRGLKVLDSLSENERSGELPESLNL 388

DB 326 TGEIPLEIGNEGLELYLENQVLTITPYE-LSTLKNLSKDLISNALTGPIPLGFOYL 384

QY 389 SASLTLTDLSSNFGPILPNLCNPNKTLQELYLQNNGTGKIPPLSLSCSLVLSLHS 448

DB 385 -RGLFMLQLFQNSLSGTIPPL--GWSYDLVWLDMSDNLHSGRIPSYLCLHSNNITLNLG 441

QY 449 FNYLSGTIPSSLSGLSKLDKLMNLBGEIPQELMYVYKTLLETLIDFNDLGEIPSGL 508

DB 442 TNNLSGNIPGTITCTVQLRLRNNLVGRFPENLCKQVNVTAIELGQRFPGSIPREV 501

QY 509 SNCTNLNWSLNNRITGEIPKWLLENLAILKLNNNSFGNIPDELGDCLSLIWLNL 568

DB 502 GNCNALQRLQDLNCGFQELPREIGMLSQGLTLNISNKLGEVSEIFNCKMLQRLDMC 561

QY 569 TNLFGTIPAMFQSGKIAANFTAGKRYVYIKNDGKKCKCHGAGNLEPFGGISBOLNR 628

DB 562 CNPFSGTLPSEV-----GSLYOL-----ELK 583

QY 623 ISTRPCNITSRVVGHNSTPTFDNNGSMWFLDMSYMLSGYIPKEISMPYLI-LNLGH 687

DB 584 LSNNN-----LSGTIPVALGNLSRLTELQNGNLENGLPRELGLTGLQALNLISY 635

QY 688 NDISGSLPDEVGLRGLNILLDSNKLDRIPQAMSAITMLTEIDLNNNSGPIPEMGQ 747

DB 636 NKLGTIPPELSNLVLEFLILNNNNLGEIPSSFANLSLIGVFSNLSITGLPPL-- 693

QY 748 FETFPKAKLNPGLCYVPLPRC-----DPSNADGVANHQSHGRPASLAGVAMGL 801

DB 694 -RNISMSFPIGNEGLCPPLNQCITQTPAPSSQSTG-----KGGMRSSKIIAIT 742

QY 802 FSPVCIFFGLVGERMKRRKREKAELEMYAEHGNSGDRTANTNTMKLTGVKALSLNL 861

DB 743 AAVTGVSLMLTALYYLMRRPVRTVASAQDG-----QPEMSL 782

QY 862 AAFKPLRLKLTADLIQATNGRHNSLISGGFGDVYKAILKDSAAVAIKKLI--HVSQ 919

DB 783 DIYFPKEGTFQDLVAATDNDPESFVVGACGVYKAVLPAGTTTAVKKLASNHEGN 842

QY 920 G--DEEVAEMETIKIGHNVLPLLYGCKYGDERTLVNEVMKYGSLEVDVLODPKGGV 976

DB 843 NNNVDSFRAEITLGNIRKNIYKLGFCNNGSMLLYEYMPKSLGELIHPD---SC 899

QY 977 KKLSTRKRLAISARGALFLHNHNSPHIIRDMKSSNVLLBENLEARSDDGMALMSA 1036

DB 900 NLDMSKKFKIALGAOGLAHLHDDCKPRIFHRDIKSNNTLLDDKFAHAGDGLAVYD- 958

QY 1037 MDTHLSVSTIAGTPGVPPPYQSFQSTGVDYSGVVLLELTGKRPDSDPFQDNIL 1096

DB 959 MPEKSMASALAGSYIAPAYATMKTEKSDYISGVVLELITLTKAVQPIPDQ--GVV 1017

QY 1097 VGVWKQAK--LRISVDFDELKEDPALEIHLQHLKVAVACLDPRAMRPTMVMAM 1154

DB 1018 VNNVRSITRRDALLSGVLDARLTLEDBRIYSHMLTYIKIALCTSVSPVAPBMRQVILM 1077

QY 1155 FKEIQAGSG 1163

DB 1078 LIESERSEG 1086



Herbicide; plant; agriculture; herbicide.

Arabidopsis thaliana.

MO200210210-A2.

07-FEB-2002.

28-AUG-2001; 2001WO-EP09892.

28-AUG-2001; 2001WO-EP09892.

(FARB) PAYER AG.

Tietjen K, Weidler M;

WPI; 2002-269010/31.

Identifying plant target proteins for herbicidally active compounds, comprising aligning and comparing nucleic acid or amino acid sequences from plant with nucleic acid or amino acid sequences from non-plant organisms -

Claim 5; SEQ ID NO 249; 2elpp + Sequence listing; English.

The invention relates to identifying target proteins (AB99790-AB994016) for herbicidally active compounds, comprising aligning and comparing nucleic acid or amino acid sequences from plant with nucleic acid or amino acid sequences from non-plant organisms using suitable search parameters, where plant sequences having an E-value greater by a factor of 3 than the E-value of most similar non-plant sequences are selected. The polypeptides or nucleic acids encoding them are useful for identifying modulators. The identified modulators are useful as herbicides.

Sequence 1133 AA;

Query Match 22.1%; Score 1367; DB 23; Length 1133;

Best Local Similarity 32.9%; Pred. No. 9.8e-95;

Matches 398; Conservative 179; Mismatches 455; Indels 176; Gaps 36;

8 PLSTVTLFFSFSSLSFQASPSQSLYREIHQISFKVDLPDKNLLPDMSS-NKNPCTFD 65

7 FLAIVILCSFSLIV-----RSINBGRVLEFRKFLANDSNGYLAAMNQLDSNPKWT 59

66 GVTCDKDD-VTSIDLSKPLNVGSFSAVSSLSLTGLSLSFLNSHNGSV-SGFKCSAS 123

60 GIACHTLRITVSDVLDNGMLS---GTLSPILCKHGRKLANVSTNFIISGPIDSLCRS 116

124 LTLSDLSRNSLSG-----PVTTLTSL-----GSCSGIKPLNVSNLTD 161

117 LEVLIDLCIRFGVPIPIQLTMIITLKLKLYICENYLGSIPIQINLSISQELVITISNNL- 175

162 FPGKVGGL-KINSLLEVLDLSANSISGANVGVLSDCGGLKHLAISGNKISGV--DV 218

176 -TGVIPPSMAKRLQIIIRAGKNGSG--VTPSEIS-GCESLKVIGLANNLEGLSPQL 231

219 SRCVNLEFLVSSNNFSTGI-PFLGDCSALQHLDISGNKLSGDFSAISTCTELKILNIS 277

232 EKIQNTLDILMQRSLSGEIPSVGNISREVLALHENYFTGSIPIREICKTKMKRLYLY 291

278 SNOFVGPIPL--PLKSLQYLSLAENKFTGERP-DPLSACDTLGLDLSGNHFGAVVP 334

292 TNOJTGIEIPEIGNLDIAEIDSENOQTGPIKEP--GHILNKLHLFENILQPIPR 349

335 FFGSCSLLELSLSSNNFSGELPMDTLTKMRGKVLDSFNEPSGELPSLTNLSASLLT 394

350 ELGBLTLEKLDLSINRLNGTIPOE-LQFLPYVDLQFLPNQLEKIP-PLIGFYNFESV 407

395 LDISNNFSGPIIPNLCONPKNTIQELYLQNGFTGKIPTTSLNGSELVSIHLSFYVLSG 454

408 LDMANSLSGPIPAHFCR--FQTLILSLSGNLSGNIPDLTKCSLTETMLGMDQLTG 465

QY 455 TIPSLSGLSKLDKDKLMLMLEGIPQELMVYKLTETLLIDENDLTGEIPSGLSNCTNL 514

DB 466 SLPIELFNQNLTALEHQLWLSGNISADGLKINLERLANNNTGELPEIGULTKI 525

QY 515 NWISLNNRLTGEIPKWIIGLEMLAILKLSNNSPSGNIPELDCRSLLWLDNTLFPNG 574

DB 526 VGRNISSNQLTGHIPKELGSCVTIQRDLISGNKFSGVIAQELQVLELIRLSDRLTG 585

QY 575 TIPAMFKOSGKIAMNFIAGKRYIYINDGMKECHQAGKLEFQGRSRQLRLSTRNP 634

DB 586 EIP-----HSFGDL-----TR-- 596

QY 635 CNISRYVGGHTSPTFDNNGSMFLDMSYMLSGYIPKEIGSMPLYFI-LNLGNDISGS 693

DB 597 -----LMELQCGNLSSENITVEIGKLTSLQISINISHNLSGT 635

QY 694 IPDEVGDLRGNTLIDLSNNKLDGRIPOAMGALTMLTIDLSNNLSGPIPEMGOFETFP 753

DB 636 IPDSLGNLOMLEIYILNDNKLSGEIPASIGLMSLLCNISNNNLVGTVPDTAVFQRMDS 695

QY 754 AKFLANPGLCGYPLPRCDP--SNADGYAHHQSHGRBPASIA-----GSVAMGLFSFV 805

DB 636 SNFAGNHGLCNSQSHCCPLVPHSDSKLMLINSQOKLITTCIYIGSV---FLITFL 752

QY 806 CIFGLILVGEREMKRRKKEALEMVAEGHNSGDRJANTNTMKLTGVKALSTINLAPE 865

DB 753 GLCWT-----KREPAFAVALE-----DQTKPDV-----MDSY 781

QY 866 KPLRLKTFADLIQATNGHNSLIGSGGFQDYVYKILKDSAVAIKKLIVHSGG---DR 922

DB 782 FPKGFYQGLVADATRNPSSEVVLGRGACGVYVAEMSGGVIAVKKL-NRGRGASSDN 840

QY 923 EMAMEETIGIKHNNVPLDGYCYGDERLLVNEVMKYGSLLEDVLPKKGAVYKLTST 982

DB 841 SFAESTLIGKIRHNIYKLYGFCYHONSLLIYVYMKSGSLGEQLQGEKXCL-LDWNA 899

QY 983 RRKIAISARGLAFLAHNCSPIHIDMKSSNVLDENLEPVSFPGMARLMSADHTLS 1042

DB 900 RRIALGAHEGICYLHDCRPQIVHRDIKSNLILDERFOAHVGFGLAKLID-LSYSKS 958

QY 1043 VSTLAGTPEVPEYVQSRCTGKDVYSYGVILLELTGKRPDTPSGFQGNLVGWYKQ 1102

DB 959 MSAVASGYIAPEYAVYTWKTEKDIYSFGVALLLELTGKRPVPLDGG-GLVNVWR 1017

QY 1103 HAKLRIS--DVEPDELKMDPALEIELQHLKVAACLDDRAMRPTVQVNAFKEIOA 1160

DB 1018 STNNMPTIEMFPAKLDPTDKRTVHMSLVLTALFCTNSPASRPTMREVVAMITEARG 1077

QY 1161 GSGIDSQS 1168

DB 1078 SSSLSSS 1085

RESULT 15

ABB91986

ID ABB91986 standard; Protein; 1124 AA.

XX ABB91986;

XX 31-MAY-2002 (first entry)

XX Herbicidally active polypeptide SEQ ID NO 1197.

XX Herbicide; plant; agriculture; herbicide.

XX Arabidopsis thaliana.

XX MO200210210-A2.

XX 07-FEB-2002.

XX 28-AUG-2001; 2001WO-EP09892.

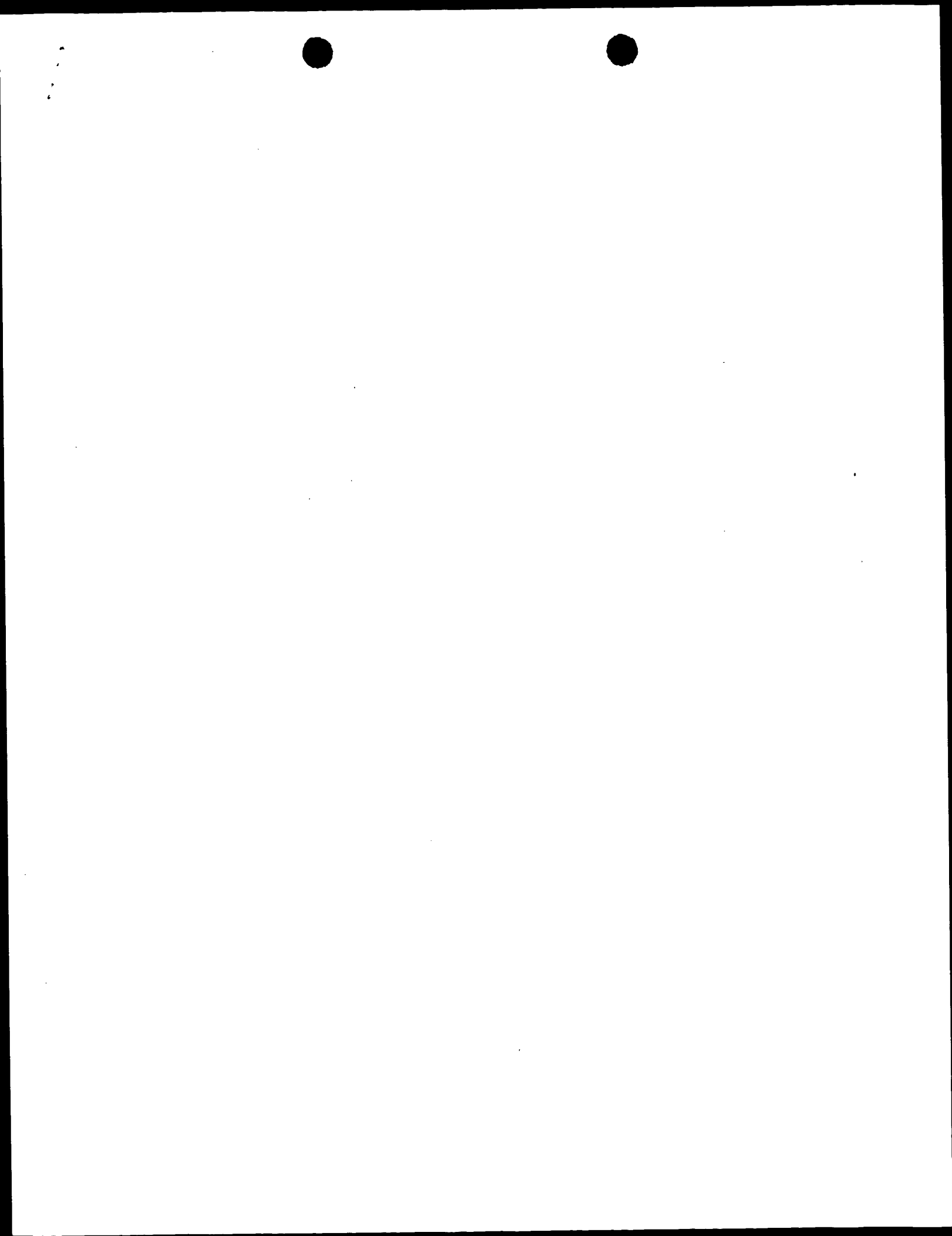
586 -KIAANFIAGKRYVYI

[illegible]

```

      1100 UNISVDFEELMNA-EDEALEIELEDDHUKVAVACEDBRAMRRPTNWQWAMFKE 1157
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1049 SLTSELDPYLTKVEDDVIINHMITVTKIAVLCTYKSSPSDRPTMEEVLMLE 1101

```





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 18:02:09 ; Search time 2.85851 Seconds

(without alignments)  
4469.091 Million cell updates/sec

Title: US-09-823-394-2\_COPY\_588\_649

Perfect score: 332

Sequence: 1 AANFIAKRVYIKNDGMRK.....STRNFCNITSRVYGGHTSPT 62

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_21:\*

1: sp\_archaea:\*

# SUMMARIES

Result No. Score Query Match length DB ID Description

Result No.	Score	Query Match length	DB ID	Description
1	332	100.0	1196 10 022476	022476 arabidopsis
2	140.5	42.3	1121 10 094273	094273 oryza sativ
3	128	38.6	1166 10 092WC8	092WC8 arabidopsis
4	124	37.3	1166 10 092ARF3	092ARF3 capsella ru
5	116	34.9	1164 10 091LJF3	091LJF3 arabidopsis
6	115	34.6	1192 10 09ARC8	09ARC8 lycopersico
7	99.5	30.0	1110 10 094LN2	094LN2 oryza sativ
8	94.5	28.5	1143 10 092S59	092S59 arabidopsis
9	65	19.6	322 2 P94168	P94168 actinobacill
10	63.5	19.1	322 2 P94168	P94168 bruceella me
11	63.5	19.1	1436 5 096158	096158 plasmodium
12	63	19.0	1436 5 096158	096158 pyrococcus
13	62.5	18.8	295 10 09MBG8	09MBG8 arabidopsis
14	62	18.7	871 17 059245	059245 pyrococcus
15	61	18.4	1592 5 09GCU17	09GCU17 lucilia cup
16	60.5	18.2	341 2 085214	085214 pseudomonas

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match length	DB ID	Description
17	60.5	18.2	466 2 0916M4	0916M4 pseudomonas
18	60.5	18.2	725 5 09VEG6	09VEG6 dirosophila
19	60.5	18.2	1341 5 09VM09	09VM09 dirosophila
20	60	18.1	1227 16 0820R1	0820R1 anabaena sp
21	60	18.1	3078 5 026031	026031 plasmodium
22	59.5	17.9	1195 5 09N343	09N343 caenorhabdi
23	59	17.8	333 16 09CHJ5	09CHJ5 lactococcus
24	59	17.8	408 16 0924S5	0924S5 salmoneilla
25	58.5	17.6	385 10 09FKX4	09FKX4 arabidopsis
26	58.5	17.6	776 17 027582	027582 methanobact
27	58.5	17.6	914 12 085427	085427 rat cytoleg
28	58	17.5	275 17 087VD3	087VD3 methanopyru
29	58	17.5	385 11 087VD3	087VD3 mus musculu
30	58	17.5	397 11 035325	035325 mus musculu
31	58	17.5	397 11 035118	035118 mus musculu
32	58	17.5	397 11 035118	035118 mus musculu
33	57.5	17.3	110 5 095X53	095X53 caenorhabdi
34	57.5	17.3	485 10 024665	024665 porphyra pu
35	57.5	17.3	523 10 09S7T6	09S7T6 arabidopsis
36	57.5	17.3	539 4 087BA7	087BA7 homo sapien
37	57.5	17.3	635 4 087BA7	087BA7 homo sapien
38	57.5	17.3	636 16 092N98	092N98 rhizobium m
39	57.5	17.3	733 11 091XW0	091XW0 rattus norv
40	57.5	17.3	842 10 091UJ2	091UJ2 arabidopsis
41	57	17.2	107 16 092GN6	092GN6 rickettsia
42	57	17.2	126 6 09GMD0	09GMD0 macaca fasc
43	57	17.2	300 16 08UAK1	08UAK1 agrobacteri
44	57	17.2	386 4 0915R3	0915R3 homo sapien
45	57	17.2	1488 16 P96204	P96204 mycobacteri

## ALIGNMENTS

### RESULT 1

ID	Score	Query Match length	DB ID	Description
022476	332	100.0	1196 10 022476	022476 arabidopsis
AC	022476	100.0	1196 10 022476	022476 arabidopsis
DT	01-JUN-1998	100.0	1196 10 022476	022476 arabidopsis
DT	01-JUN-1998	100.0	1196 10 022476	022476 arabidopsis
DT	01-JUN-2002	100.0	1196 10 022476	022476 arabidopsis
DE	Brassinosteroid insensitive 1 (Br1)	100.0	1196 10 022476	022476 arabidopsis
DE	Brassinosteroid insensitive 1 (Br1)	100.0	1196 10 022476	022476 arabidopsis
DN	BR1 OR F23K16.30 OR AT4G39400	100.0	1196 10 022476	022476 arabidopsis
OS	Arabidopsis thaliana (Mouse-ear cress)	100.0	1196 10 022476	022476 arabidopsis
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	100.0	1196 10 022476	022476 arabidopsis
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;	100.0	1196 10 022476	022476 arabidopsis
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.	100.0	1196 10 022476	022476 arabidopsis
OK	NCBI_TaxID=3702;	100.0	1196 10 022476	022476 arabidopsis
RP	SEQUENCE FROM N.A.	100.0	1196 10 022476	022476 arabidopsis
RP	SEQUENCE FROM N.A.	100.0	1196 10 022476	022476 arabidopsis
RC	STRAIN=CV. COL-O;	100.0	1196 10 022476	022476 arabidopsis
RC	MEDLINE=97442355; PubMed=9298904;	100.0	1196 10 022476	022476 arabidopsis
RT	Li U., Chory J.;	100.0	1196 10 022476	022476 arabidopsis
RT	"A putative leucine-rich repeat receptor kinase involved in	100.0	1196 10 022476	022476 arabidopsis
RT	brassinosteroid signal transduction.";	100.0	1196 10 022476	022476 arabidopsis
RL	Cell 90:929-938(1997).	100.0	1196 10 022476	022476 arabidopsis
RL	[2]	100.0	1196 10 022476	022476 arabidopsis
RA	SEQUENCE FROM N.A.	100.0	1196 10 022476	022476 arabidopsis
RA	Beyan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,	100.0	1196 10 022476	022476 arabidopsis
RA	Mayer K.F.X., Lemcke K., Scheller C.;	100.0	1196 10 022476	022476 arabidopsis
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.	100.0	1196 10 022476	022476 arabidopsis
RL	[3]	100.0	1196 10 022476	022476 arabidopsis
RP	SEQUENCE FROM N.A.	100.0	1196 10 022476	022476 arabidopsis
RP	EU Arabidopsis sequencing project;	100.0	1196 10 022476	022476 arabidopsis
RL	Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.	100.0	1196 10 022476	022476 arabidopsis
RP	SEQUENCE FROM N.A.	100.0	1196 10 022476	022476 arabidopsis
RP	Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.;	100.0	1196 10 022476	022476 arabidopsis
RP	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	100.0	1196 10 022476	022476 arabidopsis
RP	[5]	100.0	1196 10 022476	022476 arabidopsis
RP	SEQUENCE FROM N.A.	100.0	1196 10 022476	022476 arabidopsis
RA	EU Arabidopsis sequencing project;	100.0	1196 10 022476	022476 arabidopsis

Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL; AF017056; AAC49810.1; -  
 DR EMBL; AL078620; CAB4675.1; -  
 DR EMBL; AL161595; CAB80603.1; -  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR003592; LRR\_out.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR InterPro; IPR004040; STY\_pkinase.  
 DR Pfam; PF00560; LRR; 20.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR00019; LEURICHRPT.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00370; LRR; 15.  
 DR SMART; SM00221; STYK; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00101; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR ATP-binding; Serine/threonine-protein kinase; Transferase.  
 KX SEQUENCE 1196 AA; 130542 MW; CTFBAC21294E600 CRC64;  
 Query Match 100.0%; Score 332; DB 10; Length 1196;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-33;  
 Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 AANFIAGKRYVYIKNDGKKECHGAGNLEFQGISRSQLNRLSTRNPNCTTSRYGHTS 60  
 Db 588 AANFIAGKRYVYIKNDGKKECHGAGNLEFQGISRSQLNRLSTRNPNCTTSRYGHTS 647  
 Oy 61 PT 62  
 Db 648 PT 649  
 RESULT 2  
 Q942F3 PRELIMINARY; PRT; 1121 AA.  
 ID Q942F3  
 AC Q942F3;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Putative brassinosteroid-insensitive protein BR11.  
 GN P0480C01.13.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzeae; Oryza.  
 NC NCB1\_TaxID=4530;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.,  
 RA "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
 RT clone: PD48C01.";  
 RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AP003453; BAB68053.1; -  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR Pfam; PF00560; LRR; 19.  
 DR Pfam; PF00069; pkinase; 1.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; UNKNOWN\_1.  
 DR PROSITE; PS00101; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; UNKNOWN\_1.  
 KW ATP-binding; Transferase.  
 SO SEQUENCE 1121 AA; 120180 MW; F71A49B45E0E2D09 CRC64;  
 Query Match 42.3%; Score 140.5; DB 10; Length 1121;  
 Best Local Similarity 46.8%; Pred. No. 4.5e-09;  
 Matches 27; Conservative 11; Mismatches 19; Indels 1; Gaps 1;

Oy 5 IAGKRYVYIKNDGKKECHGAGNLEFQGISRSQLNRLSTRNPNCTTSRYGHTSPT 62  
 Db 520 IAGKRYVYIKNDGKKECHGAGNLEFQGISRSQLNRLSTRNPNCTTSRYGHTSPT 576  
 RESULT 3  
 Q9ZWC8 PRELIMINARY; PRT; 1166 AA.  
 ID Q9ZWC8  
 AC Q9ZWC8;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE F20N2.4.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsids.  
 NC NCB1\_TaxID=3702;  
 RP SEQUENCE FROM N.A.  
 RA Khan S., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Kim C.,  
 RA Khan P., Altafi H., Bel O., Chin C., Chio J., Choi E., Conn L.,  
 RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,  
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharasy N., Nguyen M.,  
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,  
 RA Toriumi M., Vayberg M., Yu G., Federspiel N.A., Theologis A.,  
 RA Ecker J.R.;  
 RA "Genomic sequence for Arabidopsis thaliana BAC F20N2 from chromosome  
 I.1.";  
 RT Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RA Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RA Submitted (AUG-1998) to the EMBL/Genbank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Shin P.P., Buehler E.E., Dunn P.P., Feng J.J., Kim C.C., Li Y.Y.,  
 RA Walker M.W., Altafi H.H., Araujo R.R., Conn L.L., Conway A.A.B.,  
 RA Gonzalez A.A., Hansen N.N.F., Hutzar L.L., Kremenetskaia I.I.,  
 RA Lenz C.C., Li J.J., Liu S.S., Luros S.S., Rowley D.D., Schwartz J.J.,  
 RA Toriumi M.M., Vayotskaya V.V., Yu G.G., Davis R.R.W.,  
 RA Toriumi M.M., Theologis A.A., Ecker J.R.;  
 RA Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RA Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RA Cheuk R., Shin P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
 RA Khan S., Kim C., Altafi H., Bel B., Chin C., Chio J., Choi E.,  
 RA Khan L., Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B.,  
 RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharasy N.,  
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
 RA Thaveri A., Toriumi M., Vayberg M., Yu G., Davis R., Federspiel N.,  
 RA Theologis A., Ecker J.;  
 RA Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL; AC002328; AAF79510.1; -  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR003592; LRR\_out.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR InterPro; IPR004040; STY\_pkinase.  
 DR InterPro; IPR001245; TYR\_pkinase.  
 DR Pfam; PF00560; LRR; 21.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR00019; LEURICHRPT.  
 DR PRINTS; PR00109; TYRKINASE.

DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00370; LRR; 16.  
 DR SMART: SM00221; STYKc; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.  
 DR PROSITE; PS00108; PROTEIN KINASE ST; 1.  
 DR ATP-binding; Serine/threonine-protein kinase; Transferase.  
 KM euroidae II; Brassicales; Brassicaceae; Capsella.  
 SQ SEQUENCE 1166 AA; 127423 MW; 8C4D9231A466A7 CRC64;

Query Match 38.6%; Score 128; DB 10; Length 1166;  
 Best Local Similarity 43.1%; Pred. No. 1.9e-07;  
 Matches 25; Conservative 15; Mismatches 16; Indels 2; Gaps 2;

OY 5 IAGKRYVYIKNDGKKECHGAGNLEFGIRSEQLNRLSTRNPNCTISRYVGHTSPT 62  
 Db 579 VSGKQFAFVFNNEG-GTDCRGAGGLVEFGIRARLERLHPMVHSCPAT-RISYGMTWT 634

RESULT 4  
 OY 09ARF3 PRELIMINARY; PRT; 1166 AA.  
 AC 09ARF3;  
 DT 01-JUN-2001 (TREMblrel. 17, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Hypothetical 127.3 kDa protein.  
 OS Capsella rubella.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC euroidae II; Brassicales; Brassicaceae; Capsella.  
 OC NCBI\_TaxID=81985;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21178822; PubMed=11283350;  
 RA Rosberg M., Theres K., Acarkan A., Herrero R., Schmitt T.,  
 RA Schumacher K., Schmitz G., Schmidt R.;  
 RT "Comparative sequence analysis reveals extensive microcolinearity in  
 RT the lateral suppressor regions of the tomato, Arabidopsis and Capsella  
 RT genomes.";  
 RL Plant Cell 13:979-988(2001).  
 CC -1. SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL; AJ030349; CAC36390.1; -  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00560; LRR; 18.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR00109; LEURICHRPT.  
 DR PRODOM; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00220; S\_TKc; 1.  
 DR SMART; SM00219; TYKc; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.  
 DR PROSITE; PS00108; PROTEIN KINASE ST; 1.  
 KM ATP-binding; Hypothetical protein; Serine/threonine-protein kinase;  
 Transferase.  
 SQ SEQUENCE 1166 AA; 127261 MW; DAD222B4C5647A0 CRC64;

Query Match 37.3%; Score 124; DB 10; Length 1166;  
 Best Local Similarity 41.4%; Pred. No. 6e-07;  
 Matches 24; Conservative 15; Mismatches 17; Indels 2; Gaps 2;

OY 5 IAGKRYVYIKNDGKKECHGAGNLEFGIRSEQLNRLSTRNPNCTISRYVGHTSPT 62  
 Db 579 VSGKQFAFVFNNEG-GTDCRGAGGLVEFGIRARLERLHPMVHSCPAT-RISYGMTWT 634

RESULT 5  
 OY 09LJF3 PRELIMINARY; PRT; 1164 AA.  
 AC 09LJF3;

DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Receptor protein kinase.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC euroidae II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLUMBIA;  
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLUMBIA;  
 RA MEDLINE=20363099; PubMed=10907853;  
 RA Nakamura Y.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.  
 RT Sequence features of the regions of 4,251,695 bp covered by ninety pl,  
 RT TAC and BAC clones.";  
 RL DNA Res. 7:217-221(2000).  
 CC -1. SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL; AP000603; BAB01743.1; -  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR003592; LRR\_out.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR InterPro; IPR004040; STY\_pkinase.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00560; LRR; 20.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR00109; LEURICHRPT.  
 DR PRODOM; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00370; LRR; 14.  
 DR SMART; SM00221; STYKc; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.  
 DR PROSITE; PS00108; PROTEIN KINASE ST; 1.  
 KM ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
 SQ SEQUENCE 1164 AA; 126660 MW; 79380581D400EEEC CRC64;

Query Match 34.9%; Score 116; DB 10; Length 1164;  
 Best Local Similarity 40.0%; Pred. No. 6.3e-06;  
 Matches 22; Conservative 15; Mismatches 16; Indels 2; Gaps 2;

OY 5 IAGKRYVYIKNDGKKECHGAGNLEFGIRSEQLNRLSTRNPNCTISRYVGHT 59  
 Db 579 VSGKQFAFVFNNEG-GTDCRGAGGLVEFGIRARLERLHPMVHSCPAT-RISYGMT 631

RESULT 6  
 ID 09ARC8 PRELIMINARY; PRT; 1192 AA.  
 AC 09ARC8;  
 DT 01-JUN-2001 (TREMblrel. 17, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Hypothetical 129.9 kDa protein.  
 OS Lycopersicon esculentum (Tomato).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; eusterids I; Solanales; Solanaceae; Solanum.  
 OC NCBI\_TaxID=4081;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21178822; PubMed=11283350;  
 RA Rosberg M., Theres K., Acarkan A., Herrero R., Schmitt T.,  
 RA Schumacher K., Schmitz G., Schmidt R.;  
 RT "Comparative sequence analysis reveals extensive microcolinearity in  
 RT the lateral suppressor regions of the tomato, Arabidopsis and Capsella

```

RT genomes."
CC Plant Cell 13:979-988(2001).
-1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AJ303345; CAC36401.1; -.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00560; LRR; 21.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00019; LEURICHRPT.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TYKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS50011; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN KINASE_ST; 1.
KW ATP-binding; Hypothetical protein; Serine/threonine-protein kinase;
Transferrase.
SEQUENCE 1192 AA; 129941 MW; BDEICDAP8930886 CRC64;

Query Match 34.6%; Score 115; DB 10; Length 1192;
Best Local Similarity 43.9%; Pred. No. 8.7e-06;
Matches 25; Conservative 12; Mismatches 18; Indels 2; Gaps 2;

6 AGKRVYIKNDGKKECHGAGNLEFGGIRSEQLNRLSTNPNCTSRVYGGTSTPT 62
Db 609 SGKQPAFPVANG-GTECHGAGVGFEGIRERLALIPVHFCST-RIVSGRTMYT 663

RESULT 7
094LN2 PRELIMINARY; PRT; 1110 AA.
AC 094LN2:
DT 01-DEC-2001 (TRENBLREL. 19, Created)
DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE Putative receptor protein kinase.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Euphorbiaceae; Oryzaeae; Oryza.
OC NCBI_TaxID=4530;
OK [1]
RN SOURCE FROM N.A.
RP Wing R.A., Fritsch D., Presting G., Wood T., Yu Y., Soderlund C.,
RA Kim H., Rambo T., Henry D., Simmons J.;
RA Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AC078891; AK5254.1; -.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000504; RNA_rec_mot.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00560; LRR; 19.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; UNKNOWN_1.
DR PROSITE; PS50011; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN KINASE_ST; UNKNOWN_1.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
KW ATP-binding; Kinase; Receptor; Serine/threonine-protein kinase;
Transferrase.
SEQUENCE 1110 AA; 118109 MW; B6723380BC0A8E9A CRC64;

Query Match 30.0%; Score 99.5; DB 10; Length 1110;
Best Local Similarity 37.5%; Pred. No. 0.00076;
Matches 21; Conservative 11; Mismatches 21; Indels 3; Gaps 2;

2 ANFAGKRVYIKNDGKKECHGAGNLEFGGIRSEQLNRLSTNPNCTSRVYGG 57
Db 530 SGILSGNTLAFVNVG--NSCKGVGILFPAIGIRPERLLQVPTKSCDPT-RLVSG 582

```

```

RESULT 8
092PS9 PRELIMINARY; PRT; 1143 AA.
AC 092PS9:
DT 01-MAY-1999 (TRENBLREL. 10, Created)
DT 01-MAY-1999 (TRENBLREL. 10, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE Putative receptor protein kinase.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Euposids II; Brassicales; Brassicaceae; Arabidopsids.
OK NCBI_TaxID=3702;
RN [1]
RN SOURCE FROM N.A.
RP STRAIN=CV. COLUMBIA;
RC MEDLINE=20083487; PubMed=10617197;
RX Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buehl C.R., Ketchum K.A., Lee J.J., Ronning C.M., Xoo H., Moffat K.S.,
RA Cronin L.A., Shen M., Vanden S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carreira A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhagen G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana."
RL Nature 402:761-768 (1999).
RN [2]
RN SOURCE FROM N.A.
RP STRAIN=CV. COLUMBIA;
RC Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RL [3]
RN SOURCE FROM N.A.
RP Yamada K., Banb J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.U.,
RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
RA Carrinci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamaya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinzaki K.,
RA Davis R.W., Ecker J.R., Theologis A.;
RT "Full length cDNA of gene At2G01950 (GI:15226381).";
RL Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AC06532; AAD20088.1; -.
DR EMBL; AY074313; AAL67010.1; -.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003592; LRR_out.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR004040; STY_pkinase.
DR Pfam; PF00560; LRR; 21.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00019; LEURICHRPT.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00370; LRR; 18.
DR SMART; SM00221; STYKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS50011; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN KINASE_ST; 1.
KW ATP-binding; Kinase; Receptor; Serine/threonine-protein kinase;
Transferrase.
SEQUENCE 1143 AA; 125675 MW; 7D1C8493F27A94E CRC64;

Query Match 28.5%; Score 94.5; DB 10; Length 1143;
Best Local Similarity 33.9%; Pred. No. 0.0034;
Matches 19; Conservative 13; Mismatches 21; Indels 3; Gaps 2;

2 ANFAGKRVYIKNDGKKECHGAGNLEFGGIRSEQLNRLSTNPNCTSRVYGG 57

```

DB 549 SGLSGNTWAFVRNVG--NSCKGVGLVFPISIRPERLLQIPSLKSCDFT-RMTSG 601

# RESULT 9

ID P94168 PRELIMINARY; PRT; 322 AA.

DT 01-MAY-1997 (TREMBlrel. 03, Created)  
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Mutarotase.

GN GALT.

OS Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;

OC Actinobacillus.

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

OC NCBI\_TaxID=715;

QY 16 DOMKKECHGAGNLEFGIRSEQLNRSTNPNCTIRYVGHGS 60  
DB 241 DGIYVAGRGAGHV--SGDEADITERYASRIPVVASRTYGGRTA 282

# RESULT 11

ID O96158 PRELIMINARY; PRT; 1436 AA.

DT 01-MAY-1999 (TREMBlrel. 10, Created)

DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE Hypothetical 164.8 kDa protein.

GN PF0285C.

OS Plasmodium falciparum.

OC Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

OC NCBI\_TaxID=5833;

QY 8 KRYVYIKNDG--MKKECHGAGNLEFGIRSEQ-----LNRLSTRPCN-----ITSRV 54  
DB 183 KAVSFKNDGDNKSKNTH--DNLMFLKNIRSKSNLIVNRKITHVNVNIVSGMTNKV 240

# RESULT 12

ID O9V168 PRELIMINARY; PRT; 916 AA.

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Hypothetical protein PAB0383.

GN PAB0383.

OS Pyrococcus abyssi.

OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

OC NCBI\_TaxID=29292;

DR SMART; SM00305; HintC; 1.  
 DR SMART; SM00306; HintC; 1.  
 DR PROSITE; PS00881; PROTEIN\_SPLICING; UNKNOWN\_1.  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 916 AA; 10245 MW; 32DB1F28B85155A CRC64;

Query Match 19.0%; Score 63; DB 17; Length 916;  
 Best Local Similarity 39.1%; Pred. No. 28;  
 Matches 18; Conservative 5; Mismatches 9; Indels 14; Gaps 3;

QY 10 VYIKNDGKMKKE-----CHGAGNLT-----EFGG--IRSEQLNR 41  
 DB 821 YVLAGEGAMKETFGSTCHGAGVLSRKAATROYRGRIRIOLNLR 866

RESULT 13  
 Q9MBG8 PRELIMINARY; PRT; 295 AA.  
 AC Q9MBG8;  
 ID 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 OS Genomic DNA, chromosome 3, TAC clone:KSK13.  
 DE Arabidopsis thaliana (Mouse-ear cross).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eurosidia II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLUMBIA;  
 RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLUMBIA;  
 RX MEDLINE=20277480; PubMed=10819329;  
 RA Nakamura Y.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence  
 features of the regions of 4,504,864 bp covered by sixty pl and TAC  
 clones";  
 RL DNA Res. 7:131-135(2000).  
 DR EMBL; AB025615; BAA95750.1; -;  
 DR InterPro; IPR005176; DUF298.  
 DR Pfam; PF03556; DUF298; 1.  
 SO SEQUENCE 295 AA; 34371 MW; AEF5F27C863388D4C CRC64;

Query Match 18.8%; Score 62.5; DB 10; Length 295;  
 Best Local Similarity 28.3%; Pred. No. 9.2; Mismatches 24; Indels 7; Gaps 1;  
 Matches 17; Conservative 12; Mismatches 24; Indels 7; Gaps 1;  
 QY 8 KRYVYIKNDGKMK-----ECHGAGNLTLEFGGIRSEQLNRSTNPNITSRVYGHTS 60  
 DB 228 KKTSTKGNDEDEDEDEVLFTQNSSSLNFKRITKTSNPFRCSSKPSIERSISQGFAS 287

RESULT 14  
 Q59245 PRELIMINARY; PRT; 871 AA.  
 AC Q59245;  
 ID 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical protein PH1602.  
 GN PH1602.  
 OS Pyrococcus horikoshii.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Pyrococcus.  
 OC NCBI\_TaxID=53953;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OT3;  
 RX MEDLINE=98344137; PubMed=9679194;

RA Kawarabayashi Y., Sawada M., Horikawa H., Halkawa Y., Hino Y.,  
 RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
 RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
 RA Funahashi T., Tanaka T., Kudo H., Yamazaki J., Kishida N., Oguchi A.,  
 RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
 RA Masuchi Y., Shizuya H., Kikuchi H.;  
 RT "Complete sequence and gene organization of the genome of a hyper-  
 thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
 RT DNA Res. 5:155-16(1998).  
 RL EMBL; AP000006; BAA30714.1; -;  
 DR InterPro; IPR003586; Hedgehog\_hinct.  
 DR InterPro; IPR003587; Hedgehog\_hinctn.  
 DR InterPro; IPR002203; Intein.  
 DR InterPro; IPR004042; Intein\_endonuc.  
 DR InterPro; IPR001233; UPP0027.  
 DR Pfam; PF01139; UPP0027; 2.  
 DR SMART; SM00305; HintC; 1.  
 DR PROSITE; PS00881; PROTEIN\_SPLICING; UNKNOWN\_1.  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 871 AA; 98124 MW; 989723B089288F4 CRC64;

Query Match 18.7%; Score 62; DB 17; Length 871;  
 Best Local Similarity 37.0%; Pred. No. 35;  
 Matches 17; Conservative 6; Mismatches 9; Indels 14; Gaps 3;  
 QY 10 VYIKNDGKMKKE-----CHGAGNLT-----EFGG--IRSEQLNR 41  
 DB 776 YVLAGEGAMKETFGSTCHGAGVLSRKAATROYRGRIRIOLNLR 821

RESULT 15  
 Q9GU17 PRELIMINARY; PRT; 1592 AA.  
 ID Q9GU17;  
 AC Q9GU17;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Chitin synthase 1 (EC 2.4.1.16).  
 OS Lucilia cuprina (Greenbottle fly) (Australian sheep blowfly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Cecidoidea; Calliphoridae; Lucilia.  
 OC NCBI\_TaxID=7375;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20453125; PubMed=1098064;  
 RA Tellam R.L., Vuocolo T., Johnson S.E., Jarney J., Pearson R.D.;  
 RT "Insect chitin synthase. cDNA sequence, gene organization and  
 expression.";  
 RL Eur. J. Biochem. 267:6025-6043(2000).  
 DR EMBL; AF221067; AAG09712.1; -;  
 DR InterPro; IPR004835; Fungi\_chitin\_syn.  
 DR Pfam; PF03142; Chitin\_synth.2; 2.  
 KW Glycosyltransferase; Transferase.  
 SO SEQUENCE 1592 AA; 180716 MW; 0A440B36D0A867 CRC64;

Query Match 18.4%; Score 61; DB 5; Length 1592;  
 Best Local Similarity 31.6%; Pred. No. 91;  
 Matches 18; Conservative 9; Mismatches 20; Indels 10; Gaps 2;  
 QY 11 VYIKNDGKMKKECHGAGNLTLEFGGIRSEQLNR-----RLSTNPNITSRVYGHT 59  
 DB 1535 VFVNVGSPSEQIYGSNNGAVNGY--EHVNEDDGNSLRITNPQVYTWGYTSNT 1589

Search completed: March 10, 2003, 18:18:43  
 Job time : 4.85851 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 18:16:35 ; Search time 0.985692 Seconds

(without alignments)  
2608.861 Million cell updates/sec

Title: US-09-823-394-2\_COPY\_588\_649

Sequence: 1 AANFLAGKRYVYIKNDGKMKK.....STRNPNITSRVYGHTSPT 62

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	61	18.4	313	1 PLDB_HAEMIN	P44800 haemophilus
2	58.5	17.6	316	1 YOUN CAEEL	P30634 caenorhabdi
3	58	17.5	342	1 PG11_PHAUV	P35314 phaeocystis
4	58	17.5	2210	1 RRPL_EBOSM	O66892 ebola virus
5	57.5	17.3	719	1 HS9A_HORSE	O99XK7 equus caball
6	57.5	17.3	728	1 HS9A_CHICK	P11501 gallus gall
7	57.5	17.3	731	1 HS9A_HUMAN	P07900 homo sapien
8	57.5	17.3	732	1 HS9A_CRICK	P46633 cricetus
9	57.5	17.3	732	1 HS9A_MOUSE	P07901 mus musculu
10	57.5	17.3	732	1 HS9A_MOUSE	O02705 sus scrofa
11	57	17.2	342	1 PG13_PHAUV	P58823 phaeocystis
12	56	16.9	257	1 DLH1_SUISO	P95862 sulfolobus
13	56	16.9	337	1 G55A_CHICK	O98892 gallus gall
14	56	16.9	3462	1 RELN_RAT	P58751 rattus norv
15	55.5	16.7	118	1 MTPN_CHICK	O91955 gallus gall
16	55.5	16.7	658	1 TKT_STRPN	P22976 streptococc
17	55	16.6	135	1 ES30_TRICO	O97391 cricetion
18	55	16.6	467	1 EM55_FUGRU	P49697 fuigu tridip
19	55	16.6	474	1 Y068_MTCGE	P47314 mycoplasma
20	55	16.6	1064	1 NFC2_MOUSE	O60591 mus musculu
21	54.5	16.4	383	1 DNAQ_LACSK	O87788 lactobacill
22	54.5	16.4	384	1 PQOE_ACICA	O06601 comamonas c
23	54.5	16.4	342	1 301D_COMTE	P58822 phaeocystis
24	54	16.3	342	1 EXL1_PHAUV	O97400 clostridium
25	54	16.3	399	1 EXL1_CLOAB	P24617 neisseria g
26	53.5	16.1	278	1 T2P2_NEIGO	P47047 saccharomyc
27	53.5	16.1	1073	1 MTR4_YEAST	P32736 ratius norv
28	53	16.0	345	1 APCM_RAT	P47927 arabidopsis
29	53	16.0	432	1 AP2_ARATH	O11125 caenorhabdi
30	53	16.0	706	1 YX14_CAEEL	O42846 schizosacch
31	52.5	15.8	112	1 RL14_SCHPO	P11927 saccharomyc
32	52.5	15.8	433	1 KARI_YEAST	O03673 saccharomyc
33	52	15.7	198	1 YMO3_YEAST	

34	52	15.7	229	1 C79B_HUMAN	P40259 homo sapien
35	52	15.7	401	1 KYB1_CHICK	O99W1 gallus gall
36	52	15.7	710	1 PECE_BOVIN	O97881 aquifex aeo
37	51.5	15.5	216	1 PUR7_AQUAE	P70960 bacillus su
38	51.5	15.5	227	1 YWMC_BACSU	O02878 archaeoglob
39	51.5	15.5	294	1 PANE_ARCFU	O87389 thizobium m
40	51.5	15.5	324	1 GLXA_RHIME	P16602 compox virtu
41	51.5	15.5	1284	1 ARTI_COMEX	P38428 saccharomyc
42	51	15.4	203	1 YMBE_YEAST	P50265 candida alb
43	51	15.4	324	1 DLH1_CANAL	O14982 homo sapien
44	51	15.4	345	1 OPCM_HUMAN	
45	51	15.4	392	1 SB11_HUMAN	O96p15 homo sapien

## ALIGNMENTS

RESULT 1	PLDB_HAEMIN	STANDARD	PRT: 313 AA.
ID	PLDB_HAEMIN		
AC	P44800:		
DT	01-NOV-1995 (Rel. 32, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DE	Probable lysophospholipase L2 (EC 3.1.1.5) (lecithinase B).		
OS	Haemophilus influenzae.		
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;		
OX	NCBI_TaxID=727;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Rd / KM20 / ATCC 51907;		
RX	MEDLINE=95350630; PubMed=7542800;		
RA	Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness B.F.,		
RA	Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,		
RA	McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,		
RA	Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,		
RA	Wedman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,		
RA	Uterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,		
RA	Pine L.D., Fritchman J.L., Fuhrman J.L., Geoghagen N.S.M.,		
RA	Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,		
RT	Venter J.C.;		
RT	"Whole-genome random sequencing and assembly of Haemophilus influenzae		
RL	Rd."		
CC	Science 269:496-512 (1995).		
CC	-1- CATALYTIC ACTIVITY: 2-lysophosphatidylcholine + H(2)O =		
CC	glycerophosphocholine + a fatty acid anion.		
CC	-1- SUBCELLULAR LOCATION: Inner membrane (By similarity).		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL: U37747; AAC22305.1; -		
DR	TIGR: H10645; -		
DR	InterPro: IPR000073; Abhydrolase.		
DR	InterPro: IPR000379; Ser_ests_site.		
DR	Pfam: PF00561; abhydrolase; 1.		
KW	Hydrolase; Lipid synthesis; Inner membrane; Complete proteome.		
SQ	SEQUENCE 313 AA; 36656 MW; 891E784FF77C28 CRC64;		
QY	Query Match	18.4%;	Score 61; DB 1; Length 313;
QY	Best Local Similarity	36.6%;	Pred. No. 1.6;
QY	Matches 15; Conservative	9;	Mismatches 15; Indels 2; Gaps 1;
Db	7 GKRYVYIKNDGKKEGAGNLLPFGIRSEBOLNLTSTNP 47		
Db	181 GERYVFGK-GAYQQAHLVNEILTFCKTRMKMKRINKNP 219		





SEQUENCE 342 AA; 37101 MW; 950F94E0DD2A39598 CRC64;

Query Match 17.5%; Score 58; DB 1; Length 342;  
Best Local Similarity 34.5%; Pred. No. 4.3;  
Matches 20; Conservative 7; Mismatches 21; Indels 10; Gaps 3;

3 NFINGKRVYIKNDGKKECHGAGNLEFQ---GISEQLNRLSTNPNCTTSRYVG 56  
Db 236 NMLBDSAVLFGSDNKTCKHLAKNSLAFLDGKVGTL-SKINLGDLENN-----NRIYG 287

## RESULT 4

RRPL\_EBOSM STANDARD; PRT; 2210 AA.

AC 066802;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE RNA-directed RNA polymerase (EC 2.7.7.48) (Large structural protein)  
CC (L protein).

OC Ebola virus (strain Sudan Maleo-79) (Ebo).

OC Viruses; ssRNA negative-strand viruses; Mononegavirales; Filoviridae;  
OC Ebola-like viruses.

OX NCBI\_TaxID=128949;

RN [1]

RP SEQUENCE FROM N.A.

RA Sanchez A., Trappler S., Nichol S.T.,  
Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE ACTIVE POLYMERASE. IT  
MAY FUNCTION IN RNA SYNTHESIS, CAPPING, AS WELL AS METHYLATION OF  
CAPS. AND POLY(A) SYNTHESIS. ALSO INVOLVED IN TRANSCRIPTIONAL RNA  
EDITING.

CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
(RNA) (N).

CC -1- SIMILARITY: PRIMARILY WITH THE N-TERMINAL HALF OF THE L PROTEINS  
OF RHABDOVIRUSES AND PARAMYXOVIRUSES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U23458; AAA7970.1; -

DR InterPro; IPR001016; Viral\_RNA\_pol\_L.

DR Pfam; PF00946; Paramyx\_RNA\_pol\_1.

DR Ktransferrase; RNA-directed RNA polymerase.

SQ SEQUENCE 2210 AA; 251255 MW; F50EBB624951AER4 CRC64;

Query Match 17.5%; Score 58; DB 1; Length 2210;  
Best Local Similarity 39.5%; Pred. No. 34;  
Matches 17; Conservative 8; Mismatches 14; Indels 4; Gaps 2;

21 ECHGAGNLEFQ---GIRSEQLNRLSTNPNCTTSRYVGHTSP 61  
Db 1832 EEEGSGALLIQRKVKPLFNTLATER--STESVIGSYTTP 1872

## RESULT 5

HS9A\_HORSE STANDARD; PRT; 719 AA.

AC 09GKX7;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Heat shock protein HSP 90-alpha (HSP 86) (Fragment).

CC Equus caballus (Horse).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Perissodactyla; Equidae; Equus.

OX NCBI\_TaxID=9796;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21155861; PubMed=11258446;

RA Pepin K., Momose F., Ishida N., Nagata K.;

RT "Molecular cloning of horse Hsp90 cDNA and its comparative analysis  
with other vertebrate Hsp90 sequences."

RL J. Vet. Med. Sci. 63:115-124(2001).

CC -1- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY  
(BY SIMILARITY).

CC -1- SUBUNIT: Homodimer (By similarity).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AB043677; BAB20777.1; -

DR HSSP; P07900; IBO.

DR InterPro; IPR003594; ATPbind\_ATPase.

DR InterPro; IPR001404; Hsp90.

DR Pfam; PF00183; HSP90\_1.

DR Pfam; PF02518; HATPase\_C\_1.

DR PRINTS; PR00775; HEATSHOCK90.

DR SMART; SM00387; HATPase\_C\_1.

DR PROSITE; PS00298; HSP90\_1.

KW Chaperone; ATP-binding; Heat shock; Phosphorylation.

FT NON\_TER 1 1

FT MOD\_RES 220 220

FT MOD\_RES 252 252

FT NON\_TER 719 719

SQ SEQUENCE 719 AA; 83098 MW; EDCB1EE439FAC44D CRC64;

Query Match 17.3%; Score 57.5; DB 1; Length 719;  
Best Local Similarity 37.0%; Pred. No. 11;  
Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;

14 KNDGKKECHGAGNLEFQIRSEQLNRL--LSTRNPNCTTSRYVG 56  
Db 259 KKGDDKKK---KKIKKXYIDGELNKTPIWTRNDPDLTNEYG 300

## RESULT 6

HS9A\_CHICK STANDARD; PRT; 728 AA.

AC P11501;

DT 01-OCT-1989 (Rel. 12, Created)

DT 01-AUG-1990 (Rel. 15, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Heat shock protein HSP 90-alpha.

CC Gallus gallus (Chicken).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
Gallus.

OX NCBI\_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Stomach, and Gizzard;

RX MEDLINE=89165846; PubMed=2923621;

RA Binart N., Chambraud B., Dumas B., Rowlands D.A., Bigogne C.,  
Levin J.M., Garnier J., Baulieu E.E., Catelli M.G.;

RT "The cDNA-derived amino acid sequence of chick heat shock protein Mr  
90,000 (HSP 90) reveals a 'DNA like' structure: potential site of  
interaction with steroid receptors."

RL Biochem. Biophys. Res. Commun. 159:140-147(1989).

RN [2]

RP SEQUENCE OF 1-85 FROM N.A.

RX MEDLINE=89345085; PubMed=2762125;  
 RA Vourc'H C., Binart N., Chambraud B., David J.P., Jerome V.,  
 RA Baillet E.E., Catelli M.G.;  
 RT "Isolation and functional analysis of chicken 90-kDa heat shock  
 RT protein gene promoter.";  
 RL Nucleic Acids Res. 17:5259-5272(1989).  
 CC -1- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY  
 CC (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 D EMBL; X07265; CAA30251.1; -;  
 D EMBL; X15028; CAA3132.1; -;  
 D PIR; S10880; HHC90.  
 DR HSP, P07900; IBYQ.  
 DR InterPro; IPR003594; ATPbind\_ATPase.  
 DR InterPro; IPR01404; Hsp90.  
 DR Pfam; PF00183; HSP90; 1.  
 DR Pfam; PF02518; HATPase\_C; 1.  
 DR PRINTS; PR00775; HEATSHOCK90.  
 DR SMART; SM00387; HATPase\_C; 1.  
 DR PROSITE; PS00298; HSP90; 1.  
 KW Chaperone; ATP-binding; Heat shock; Phosphorylation.  
 FT MOD\_RES 230 230 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 259 259 PHOSPHORYLATION (BY SIMILARITY).  
 SQ SEQUENCE 728 AA; 84059 MW; 6EC5660FAD5006CE CRC64;  
 Query Match 17.3%; Score 57.5; DB 1; Length 728;  
 Best Local Similarity 37.0%; Pred. No. 12;  
 Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;  
 Cy 14 KNIDGKKECHGAGNLEFGIRSEQLNR---LSTNPCHITSRYVG 56  
 Db 265 KKDGGKKK----KKIKKIKYIDDELNTKPIWTNPDPDINTEEG 306  
 RESULT 7  
 HS9A\_HUMAN STANDARD; PRT; 731 AA.  
 ID HS9A\_HUMAN  
 UN P07900; OSBVOS;  
 UN 01-AUG-1988 (Rel. 08, Created)  
 UN 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Heat shock protein HSP 90-alpha (HSP 86).  
 GN HSP90A OR HSP91 OR HSP90A.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RP TISSUE=Peripheral blood lymphocytes;  
 RX MEDLINE=89386066; PubMed=2780322;  
 RA Soeda E., Yokoyama K., Yamazaki M., Akaogi K., Miwa T., Imai T.;  
 RT "Nucleotide sequence of a full-length cDNA for 90 kDa heat-shock  
 RT protein from human peripheral blood lymphocytes.";  
 RL Nucleic Acids Res. 17:7108-7108(1989).  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE=91242090; PubMed=1368637;  
 RA Yamazaki M., Tashiro H., Yokoyama K., Soeda E.;  
 RT "Molecular cloning of cDNA encoding a human heat-shock protein whose  
 RT expression is induced by adenovirus type 12 E1A in HeLa cells.";  
 RL Agric. Biol. Chem. 54:3163-3170(1990).  
 RN [3]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=89343979; PubMed=2527334;  
 RA Hickey E., Brandon S.E., Smale G., Lloyd D., Weber L.A.;  
 RT "Sequence and regulation of a gene encoding a human 89-Kilodalton  
 RT heat shock protein.";  
 RL Mol. Cell. Biol. 9:2615-2626(1989).  
 RN [14]  
 RP SEQUENCE OF 1-311 FROM N.A.  
 RX MEDLINE=88056312; PubMed=2445630;  
 RA Hoffmann T., Hovemann B.;  
 RT "Cloning and nucleotide sequence of the murine hsp94 cDNA and  
 RT chromosome assignment of related sequences.";  
 RL Gene 56:29-40(1987).  
 RN [15]  
 RP SEQUENCE OF 1-311 FROM N.A.  
 RX MEDLINE=90076956; PubMed=2591742;  
 RA Walter T., Drabant B., Krebs H., Tomalak M., Heiss S.,  
 RA Benecke B.J.J.;  
 RT "Cloning and analysis of a human 86-kDa heat-shock-protein-encoding  
 RT gene.";  
 RL Gene 83:105-115(1989).  
 RN [16]  
 RP SEQUENCE OF 184-731 FROM N.A.  
 RC TISSUE=Placenta;  
 RA Straubeberg R.;  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 RN [17]  
 RP SEQUENCE OF 538-731 FROM N.A.  
 RC TISSUE=Heart;  
 RA Tanaka M., Tanaka T., Mitani Y., Yamamoto M., Wood J.N.;  
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
 RN [18]  
 RP SEQUENCE OF 1-20, AND PHOSPHORYLATION.  
 RX MEDLINE=89123325; PubMed=2492519;  
 RA Lees-Miller S., Anderson C.W.;  
 RT "Two human 90-kDa heat shock proteins are phosphorylated in vitro at  
 RT conserved serines that are phosphorylated in vitro by casein kinase  
 RT II.";  
 RL J. Biol. Chem. 264:2431-2437(1989).  
 RN [19]  
 RP PHOSPHORYLATION BY DS-DNA KINASE.  
 RX MEDLINE=9008887; PubMed=2507541;  
 RA Lees-Miller S., Anderson C.W.;  
 RT "The human double-stranded-DNA-activated protein kinase phosphorylates  
 RT the 90-kDa heat-shock protein, hsp90 alpha at two NH2-terminal  
 RT threonine residues.";  
 RL J. Biol. Chem. 264:17275-17280(1989).  
 RN [10]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 10-222.  
 RX MEDLINE=9762065; PubMed=9108479;  
 RA Streibing C.E., Russo A.A., Schneider C., Rosen N., Hartl F.U.,  
 RA Pavletich N.P.;  
 RT "Crystal structure of an Hsp90-geldanamycin complex: targeting of a  
 RT protein chaperone by an antitumor agent.";  
 RL Cell 89:239-250(1997).  
 RN [11]  
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 10-222.  
 RX MEDLINE=99034582; PubMed=9817749;  
 RA Obermann W.M., Sonderegger H., Russo A.A., Pavletich N.P., Hartl F.U.;  
 RT "In vivo function of Hsp90 is dependent on ATP binding and ATP  
 RT hydrolysis.";  
 RL J. Cell Biol. 143:901-910(1998).  
 CC -1- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY  
 CC (BY SIMILARITY).  
 CC -1- SUBUNIT: Homodimer.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way



RX MEDLINE=90033873; PubMed=2806771;  
 RA Legnoux V., Mezger V., Quelard C., Barnier J.V., Bensaude O.,  
 RA Morange M.;  
 RT "High constitutive transcription of HSP86 gene in murine embryonal  
 RT carcinoma cells";  
 RL Differentiation 41:42-48(1989).  
 RN  
 RP SEQUENCE OF 1-30.  
 RX MEDLINE=86205848; PubMed=3458168;  
 RA Ullrich S.J., Robinson E.A., Law L.W., Willingham M., Appella E.;  
 RT "A mouse tumor-specific transplantation antigen is a heat shock-  
 RT related protein";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:3121-3125(1986).  
 RN  
 RP PHOSPHORYLATION BY DS-DNA KINASE.  
 RX MEDLINE=90008887; PubMed=2507541;  
 RA Lees-Miller S., Anderson C.W.;  
 RT "The human double-stranded DNA-activated protein kinase phosphorylates  
 RT the 90-kDa heat-shock protein, hsp90 alpha at two NH2-terminal  
 RT threonine residues";  
 RL J. Biol. Chem. 264:17275-17280(1989).  
 CC  
 CC -1- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY  
 CC (BY SIMILARITY).  
 CC -1- SUBUNIT: Homodimer.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL: U04633; AAA53068.1; -;  
 DR EMBL: M36830; AAA37868.1; ALT\_SEQ.  
 DR EMBL: M57673; AAA37867.1; -;  
 DR EMBL: X16857; CAA34748.1; -;  
 DR PIR: B32848; HHMS86.  
 DR HSP: P07900; IBYO.  
 DR MG: MGI:96250; Hsp86-1.  
 DR InterPro: IPR003594; ATPbind\_ATPase.  
 DR InterPro: IPR001404; Hsp90.  
 DR Pfam: PF00183; HSP90\_1.  
 DR Pfam: PF02518; HATPase\_c\_1.  
 DR PRINTS: PR00775; HEATSHOCK90.  
 DR SMART: SM00387; HATPase\_c\_1.  
 DR PROSITE: PS00298; HSP90\_1.  
 DR Chaperone; ATP-binding; Heat shock; Phosphorylation.  
 FT INIT MET 0 0  
 FT MOD RES 4 4 PHOSPHORYLATION (BY DS-DNA KINASE).  
 FT MOD RES 6 6 PHOSPHORYLATION (BY DS-DNA KINASE).  
 FT MOD RES 230 230 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD RES 262 262 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD RES 242 245 T->A (IN REF. 2).  
 FT CONFLICT 242 245 MISSING (IN REF. 2).  
 FT CONFLICT 355 355 R->K (IN REF. 2).  
 SQ SEQUENCE 732 AA; 84656 MW; E6345DEBACOF1B CRC64;  
 Query Match 17.3%; Score 57.5; DB 1; Length 732;  
 Best Local Similarity 37.0%; Pred. No. 12;  
 Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;

DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Heat shock protein HSP 90-alpha (HSP 86).  
 GN HSP90A OR HSP90A.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OC NCBI\_TaxID=9823;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=9912303; PubMed=9931505;  
 RA Huang H.W., Lee W.C., Lin J.H., Ulan S.C., Mao S.J., Yang P.C.,  
 RA Huang T.Y., Liu Y.C.;  
 RT "Molecular cloning and characterization of porcine cDNA encoding a  
 RT 90-kDa heat shock protein and its expression following  
 RT hyperthermia";  
 RL Gene 226:307-315(1999).  
 CC  
 CC -1- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY  
 CC (BY SIMILARITY).  
 CC -1- SUBUNIT: Homodimer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- TISSUE SPECIFICITY: Preferentially expressed in pituitary gland,  
 CC brain, adrenal gland, and testis, in comparison to other tissues.  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL: U94395; AAC8718.1; -;  
 DR HSP: P07900; IBYO.  
 DR InterPro: IPR003594; ATPbind\_ATPase.  
 DR InterPro: IPR001404; Hsp90.  
 DR Pfam: PF00183; HSP90\_1.  
 DR Pfam: PF02518; HATPase\_c\_1.  
 DR PRINTS: PR00775; HEATSHOCK90.  
 DR SMART: SM00387; HATPase\_c\_1.  
 DR PROSITE: PS00298; HSP90\_1.  
 DR Chaperone; ATP-binding; Heat shock; Phosphorylation.  
 FT INIT MET 0 0  
 FT MOD RES 4 4 PHOSPHORYLATION (BY DS-DNA KINASE).  
 FT MOD RES 6 6 PHOSPHORYLATION (BY DS-DNA KINASE).  
 FT MOD RES 230 230 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD RES 262 262 PHOSPHORYLATION (BY SIMILARITY).  
 SQ SEQUENCE 732 AA; 84643 MW; 759850062F4E5490 CRC64;

Query Match 17.3%; Score 57.5; DB 1; Length 732;  
 Best Local Similarity 37.0%; Pred. No. 12;  
 Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;  
 Db 269 KJGDGDKK---KKKIKKYYIDDELINKTKPIWTRNPDITNEYG 310  
 QY 14 KNDGKKKCHGAGNLEFQGISSEQLR--LSTRNPNITSRVYG 56  
 DB 269 KJGDGDKK---KKKIKKYYIDDELINKTKPIWTRNPDITNEYG 310  
 RESULT 11  
 PG13 PHAVU STANDARD; PRT; 342 AA.  
 AC P56823;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Polylacturonase inhibitor 3 precursor (Polylacturonase-inhibiting  
 DE protein) (PGIP-2) (PGIP-3).  
 GN PGIP3.

OS Phaseolus vulgaris (Kidney bean) (French bean).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabiales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.  
 OK NCBI\_TaxID=3885;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=cv. Pinto; TISSUE=Hypocotyl;  
 RX MEDLINE=93272053; PubMed=1303801;  
 RA Toubart P., Desiderio A., Salvi G., Cervone F., Daroda L.,  
 de Lorenzo G., Bergmann C., Davylli A.G., Alberheim P.;  
 RT "Cloning and characterization of the gene encoding the  
 endopolygalacturonase-inhibiting protein (PGIP) of Phaseolus vulgaris  
 L.";  
 RL Plant J. 2:367-373(1992).  
 CC -1- FUNCTION: Inhibitor of fungal polygalacturonase. It is an  
 important factor for plant resistance to phytopathogenic fungi.  
 CC -1- SUBCELLULAR LOCATION: Cell-wall associated.  
 CC -1- TISSUE SPECIFICITY: Found in suspension-cultured cells and to a  
 lesser extent in hypocotyls, leaves and flowers.  
 CC -1- SIMILARITY: BELONGS TO THE POLYGALACTURONASE-INHIBITING PROTEIN  
 FAMILY.  
 CC -1- SIMILARITY: CONTAINS 7 LEUCINE-RICH REPEATS (LRR).  
 CC -1- CAUTION: It is uncertain whether Met-1 or Met-10 is the initiator.  
 DR InterPro: IPR001611; LRR.  
 DR Pfam: PF00560; LRR\_5.  
 DR SMART: SM00370; LRR; 5.  
 KW Signal; Glycoprotein; Cell wall; Repeat; Leucine-rich repeat.  
 FT CHAIN 1 29  
 FT REPEAT 30 342 POLYGALACTURONASE INHIBITOR 3.  
 FT REPEAT 78 104 LRR 1.  
 FT REPEAT 129 152 LRR 2.  
 FT REPEAT 153 177 LRR 3.  
 FT REPEAT 179 200 LRR 4.  
 FT REPEAT 201 225 LRR 5.  
 FT REPEAT 225 272 LRR 6.  
 FT REPEAT 272 296 LRR 7.  
 FT REPEAT 298 321 LRR 7.  
 FT DISULFID 32 62 BY SIMILARITY.  
 FT DISULFID 63 72 BY SIMILARITY.  
 FT DISULFID 310 332 BY SIMILARITY.  
 FT DISULFID 334 341 BY SIMILARITY.  
 FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 303 303 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 342 AA; 37251 MW; 980AEP1DJA28F27D CRC64;  
 Query Match 17.2%; Score 57; DB 1; Length 342;  
 Best Local Similarity 32.8%; Pred. No. 5.8;  
 Matches 19; Conservative 9; Mismatches 20; Indels 10; Gaps 3;  
 QY 3 NFIAKRYVYIKNDGKKECHGAGNLEFQ-----GIRSEQINRSTNPNCTTSRYTG 56  
 DB 236 MNLQGDASTVPSGDNTQKXIKLAKNSLDPLDERVGL-SKNIINGDLNR-----NRIYG 287  
 RESULT 12  
 ID DLH SULSO STANDARD; PRT; 257 AA.  
 AC P95862;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Putative carboxymethylglutaminylase (EC 3.1.1.45) (Dienelectone  
 hydrolase) (DLH).  
 DE SSO2087 OR C06015 OR G31\_045.  
 OS Sulfolobus solfataricus.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.  
 OK NCBI\_TaxID=2287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;

RX MEDLINE=97055432; PubMed=8999719;  
 RA Sensen C.W., Klenk H.-P., Singh R.K., Allard G., Chan C.C.-Y.,  
 Liu O.Y., Penny S.L., Young F., Schenk W.E., Gaasterland T.,  
 RA Doolittle W.F., Ragan M.A., Charlebois R.L.;  
 RT "Organizational characteristics and information content of an  
 archaeal genome: 156 kb of sequence from Sulfolobus solfataricus  
 P2.";  
 RL Mol. Microbiol. 22:175-191(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;  
 RX MEDLINE=21332296; PubMed=11427726;  
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
 RA Aweez M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,  
 RA De Moers A., Brauso G., Fletcher C., Gordon P.M.K.,  
 RA Hikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theraulot C., Tolstrup N.,  
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
 RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;  
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
 CC -1- CATALYTIC ACTIVITY: 4-carboxymethylglutamate-2-en-4-olide + H(2)O = 4  
 oxohex-2-enedioate.  
 CC -1- SIMILARITY: BELONGS TO THE DIENELECTONE HYDROLASE FAMILY.  
 CC -1- SIMILARITY: CONTAINS 7 LEUCINE-RICH REPEATS (LRR).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Y08256; CAA69498.1; -  
 DR EMBL: AE006815; AAK42256.1; -  
 DR InterPro: IPR002925; DLH.  
 DR Pfam: PF01738; DLH; 1.  
 KW Hypothetical protein; Hydrolase; Complete proteome.  
 FT ACT SITE 148 148 BY SIMILARITY.  
 FT ACT SITE 195 195 BY SIMILARITY.  
 FT ACT SITE 226 226 BY SIMILARITY.  
 SQ SEQUENCE 257 AA; 29016 MW; B2BD7FACD6126382 CRC64;  
 Query Match 16.9%; Score 56; DB 1; Length 257;  
 Best Local Similarity 32.2%; Pred. No. 5.7;  
 Matches 19; Conservative 6; Mismatches 18; Indels 16; Gaps 3;  
 QY 8 KRYVYIKNDGKKE-----ECHGAGNLEFQIRSEQINRSTNPNCTTSRYTG 61  
 DB 129 KAYEVSSQGVKIVSMGFCWGG--LAFQ-----LATEVPIDGTIVFYGRNPP 176  
 RESULT 13  
 ID G55A CHICK STANDARD; PRT; 337 AA.  
 AC Q98892;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 30-NOV-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Neurite inhibitor GP55-A precursor (OBSCAM protein gamma isoform).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OK NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Lodge A.P., Kim D.-S., Howard M.R., McNamee C.J., Smith N., Moss D.J.;  
 RT "Cloning of CEPU-8, a secreted isoform of CEPU-1, and OBSCAM cDNAs from  
 RT chick: structural diversity of IGLON family proteins.";  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]

RP SEQUENCE OF 64-337 FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Brain; PubMed=9004047;  
 RX MEDLINE=97157768; PubMed=9004047;  
 RA Wilson D.J.A., Kim D.-S., Clarke G.A., Marshall-Clarke S., Moss D.J.,  
 "A family of glycoproteins (GP55), which inhibit neurite outgrowth,  
 are members of the Ig superfamily and are related to OSCM,  
 neurotrophin, LAMP and CEP-1."  
 RT J. Cell Sci. 109:3129-3138(1996).  
 RL J. Cell Sci. 109:3129-3138(1996).  
 CC -1- FUNCTION: INHIBITS NEURITE OUTGROWTH.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- TISSUE SPECIFICITY: RESTRICTED TO THE NERVOUS SYSTEM.  
 CC -1- DEVELOPMENTAL STAGE: INCREASES DURING DEVELOPMENT FROM VERY LOW  
 CC LEVELS AT EMBRYONIC DAY 10 AND IS MOST ABUNDANT AFTER HATCHING.  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. IGLON  
 CC SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Y08170; CAB41420.1; -  
 CC InterPro: IPR003006; IG\_MHC.  
 CC InterPro: IPR003598; IG\_C2.  
 CC InterPro: IPR003600; IG\_Like.  
 CC Pfam; PF00047; IG; 3.  
 CC SMART; SM00410; IG\_Like; 1.  
 CC DR Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;  
 KW Repeat: Signal.  
 FT SIGNAL 1 20 BY SIMILARITY.  
 FT CHAIN 21 317 NEURITE INHIBITOR GP55-A (POTENTIAL).  
 FT PROPEP 318 337 REMOVED IN MATURE FORM (POTENTIAL).  
 FT DOMAIN 43 115 IG-LIKE C2-TYPE DOMAIN 1.  
 FT DOMAIN 143 201 IG-LIKE C2-TYPE DOMAIN 2.  
 FT DOMAIN 229 295 IG-LIKE C2-TYPE DOMAIN 3.  
 FT DISULFID 50 108 POTENTIAL.  
 FT DISULFID 150 194 POTENTIAL.  
 FT DISULFID 236 288 POTENTIAL.  
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 277 277 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SS SEQUENCE 337 AA; 36887 MW; BAE717551856551E CRC64;  
 QY [REDACTED] 16.9%; Score 56; DB 1; Length 337;  
 Best Local Similarity 39.3%; Pred. No. 7.7; Indels 0; Gaps 0;  
 Matches 11; Conservative 6; Mismatches 11;  
 DB 260 EGVRIKSKGLSLTFTFNVSEKDYGYT 287  
 QY 32 OGIRSEQLRLSTRNCPNTRVYGGT 59  
 DB 260 EGVRIKSKGLSLTFTFNVSEKDYGYT 287  
 RESULT 14  
 REIN RAT STANDARD; PRT; 3462 AA.  
 AC P58751;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Reelin precursor (EC 3.4.21.-).  
 GN REIN.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).

RC TISSUE=Cerebellum;  
 RA Kikkawa S., Terashima T.;  
 RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP ALTERNATIVE SPLICING.  
 RX MEDLINE=99253436; PubMed=10328932;  
 RA Lambert de Rouvroit C., Bernier B., Royaux I., de Beyerck V.,  
 Goffinet A.M.;  
 RT "Evolutionarily conserved, alternative splicing of reelin during brain  
 development."  
 RL Exp. Neurol. 156:229-238(1999).  
 CC -1- FUNCTION: Extracellular matrix serine protease that plays a role  
 CC in layering of neurons in the cerebral cortex and cerebellum.  
 CC Regulates microtubule function in neurons and neuronal migration.  
 CC Affects migration of sympathetic preganglionic neurons in the  
 CC spinal cord, where it seems to act as a barrier to neuronal  
 CC migration. Enzymatic activity is important for the modulation of  
 CC cell adhesion. Binding to the extracellular domains of lipoprotein  
 CC receptors VLDLR and ApoER2 induces tyrosine phosphorylation of  
 CC Dab1 and modulation of Tau phosphorylation (By similarity).  
 CC -1- SUBUNIT: Binds to the ectodomains of VLDLR and ApoER2 (By  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).  
 CC -1- ALTERNATIVE PRODUCTS: 3 isoforms; 1 (shown here), 2 and 3; are  
 CC produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Abundantly produced during brain ontogenesis  
 CC by the Cajal-Retzius cells and other pioneer neurons located in  
 CC the telencephalic marginal layer of the cerebellum.  
 CC external granular layer of the cerebellum.  
 CC -1- DOMAIN: The basic C-terminal region is essential for secretion (By  
 CC similarity).  
 CC -1- SIMILARITY: BELONGS TO THE REELIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 15 BNR REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB049473; BAB78470.1; -  
 CC InterPro: IPR000561; EGF-Like.  
 CC InterPro: IPR002860; GH\_BNR.  
 CC InterPro: IPR002861; Reeler.  
 CC Pfam; PF00008; EGF; 5.  
 CC Pfam; PF02012; BNR; 15.  
 CC Pfam; PF02014; Reeler; 1.  
 DR PROSITE; PS00022; EGF\_1; 7.  
 DR PROSITE; PS01186; EGF\_2; 6.  
 KW Hydrolyase; Serine protease; Developmental protein; Matrix protein;  
 KW Cell adhesion; EGF-like domain; Glycoprotein; Repeat; Signal;  
 KW Alternative splicing.  
 FT SIGNAL 1 27 POTENTIAL.  
 FT CHAIN 28 3462 REELIN.  
 FT DOMAIN 41 173 EGF-LIKE 1.  
 FT DOMAIN 672 703 EGF-LIKE 2.  
 FT DOMAIN 1031 1062 EGF-LIKE 3.  
 FT DOMAIN 1410 1443 EGF-LIKE 4.  
 FT DOMAIN 1766 1797 EGF-LIKE 5.  
 FT DOMAIN 2130 2162 EGF-LIKE 6.  
 FT DOMAIN 2479 2510 EGF-LIKE 7.  
 FT DOMAIN 2854 2885 EGF-LIKE 8.  
 FT DOMAIN 3229 3261 BNR 1.  
 FT REPEAT 594 605 BNR 2.  
 FT REPEAT 800 811 BNR 3.  
 FT REPEAT 953 964 BNR 4.  
 FT REPEAT 1158 1169 BNR 5.  
 FT REPEAT 1324 1335 BNR 6.  
 FT REPEAT 1547 1558 BNR 7.  
 FT REPEAT 1687 1698 BNR 7.

FT REPEAT 1885 1896 BNR 8.  
FT REPEAT 2044 2055 BNR 9.  
FT REPEAT 2251 2262 BNR 10.  
FT REPEAT 2400 2411 BNR 11.  
FT REPEAT 2599 2610 BNR 12.  
FT REPEAT 2779 2790 BNR 13.  
FT REPEAT 2980 2991 BNR 14.  
FT REPEAT 3364 3375 BNR 15.  
FT DOMAIN 3433 3462 ARG-RICH (BASIC).  
FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 291 291 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 307 307 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 630 630 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1268 1268 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1448 1448 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1601 1601 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1751 1751 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1922 1922 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2146 2146 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2270 2270 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2318 2318 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2570 2570 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2963 2963 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 3017 3017 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 3074 3074 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 3186 3186 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 3413 3413 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 3440 3440 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARSPLIC 3430 3431 MISSING (IN ISOFORM 2).  
FT VARSPLIC 3430 3462 MISSING (IN ISOFORM 3).  
SQ SEQUENCE 3462 AA; 387525 MW; FCC8989090535F6 CRC64;

Query Match 16.9%; Score 56; DB 1; Length 3462;  
Best Local Similarity 29.7%; Pred. No. 1e+02;  
Matches 11; Conservative 7; Mismatches 19; Indels 0; Gaps 0;

QY 10 YVYIKNDGKKECHGAGNLEFGIRSEQINLSTPN 46  
Db 374 WLFPGATVYKSGSDGNAIYFHGNEGSQLNPATTRD 410

RESULT 15  
MTPN CHICK STANDARD; PRT: 118 AA.  
AC 091955;  
DEF 01-NOV-1997 (Rel. 35, Created)  
D 01-NOV-1997 (Rel. 35, Last sequence update)  
D 15-JUN-2002 (Rel. 41, Last annotation update)  
D Myotrophin (V-1 protein) (granule cell differentiation protein).  
GN MTPN.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Tissue: lens fibers;  
RX MEDLINE=96437509; PubMed=8840185;  
RA Sawada K., Agata K., Eguchi G.;  
RT "Characterization of terminally differentiated cell state by  
categorizing cDNA clones derived from chicken lens fibers.";  
RL Int. J. Dev. Biol. 40:531-535(1996)  
CC -!- FUNCTION: POTENTIAL ROLE IN CEREBELLAR MORPHOGENESIS. MAY FUNCTION  
IN DIFFERENTIATION OF CEREBELLAR NEURONS, PARTICULARLY OF GRANULE  
CELLS (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: CONTAINS 3 ANK REPEATS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; D26326; BAA05379.1; -  
DR HSSP; P80144; 2MYO.  
DR InterPro; IPR002110; ANK.  
DR Pfam; PF00023; ank; 3.  
DR SMART; SM00248; ANK; 2.  
DR PROSITE; PS50088; ANK\_REPEAT; 2.  
DR PROSITE; PS50297; ANK\_REGION; 1.  
KW Repeat; ANK repeat.  
FT REPEAT 1 30 ANK 1.  
FT REPEAT 34 65 ANK 2.  
FT REPEAT 67 98 ANK 3.  
SQ SEQUENCE 118 AA; 12886 MW; 6097EB55C6CBD2CF CRC64;

Query Match 16.7%; Score 55.5; DB 1; Length 118;  
Best Local Similarity 30.0%; Pred. No. 2.8;  
Matches 15; Conservative 8; Mismatches 22; Indels 5; Gaps 1;

QY 16 DGMKKECHGAGN-----LLEFGIRSEQINLSTPNPCNITSRYVGHTS 60  
Db 33 EGGKRLHYAADGQLEILFLLKGNADINAPDKNITPPLSAVBEHVS 82

Search completed: March 10, 2003, 18:25:27  
Job time : 1.98569 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 18:11:50 ; Search time 1.77424 Seconds

(without alignments)  
3359.364 Million cell updates/sec

Title: US-09-823-394-2\_COPY\_588\_649

Sequence: 332 1 AANFIAGKRVYIKNDGKMK.....STRNPNITSRYVGGHTSPT 62

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	332	100.0	1196	2	T09356 brassinosteroid-1n
2	128	38.6	1166	2	P06598 protein F20N2.4 [i
3	94.5	28.5	1143	2	B84431 probable receptor
4	63.5	19.1	322	2	AD3265 asparaginase (EC 3
5	63.5	19.1	1436	2	D71618 hypothetical prote
6	63	19.0	916	2	C75175 hypothetical prote
7	62	18.7	871	2	B71039 hypothetical prote
8	61	18.4	313	2	A64084 lysophospholipase
9	60	18.1	1227	2	AE1810 WD-40 repeat prote
10	60	18.1	3078	2	T28432 variant-specific s
11	59	17.8	323	2	H86716 L-asparaginase [im
12	58.5	17.6	316	1	S15799 protein-cytosine-p
13	58.5	17.6	776	2	C69072 anaerobic ribonucel
14	58	17.5	342	2	S23764 polygalacturonase-
15	58	17.5	397	2	UB0082 GPI-linked recepto
16	57.5	17.3	523	2	G96777 hypothetical prote
17	57.5	17.3	728	1	HHCH90 heat shock protein
18	57.5	17.3	732	1	HHHU66 heat shock protein
19	57.5	17.3	733	1	HHMS86 hypothetical prote
20	57	17.2	107	2	G97835 hypothetical prote
21	57	17.2	300	2	E88327 hypothetical prote
22	57	17.2	300	2	AH2955 hypothetical prote
23	57	17.2	1488	2	C70984 probable ppe8 prot
24	56.5	17.0	907	2	T27317 USF protein homolo
25	56	16.9	257	2	S73082 hypothetical prote
26	55.5	16.7	180	2	T50530 transketolase (EC
27	55.5	16.7	180	2	T50530 transketolase (EC
28	55.5	16.7	180	2	T50530 transketolase (EC
29	55.5	16.7	658	2	F95237 transketolase [imp

RESULT 1  
T09356  
brassinosteroid-insensitive protein BRI1 - Arabidopsis thaliana  
N/Alternate names: protein F23K16.30  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 24-Nov-1999  
C/Accession: T09356  
R/By: M. J. Murphy, G. J. Ridley, P. J. Hudson, S. J. Bancroft, I. J. Mewes, H. W. J. Mayer, K. F. X.  
submitted to the Protein Sequence Database, June 1999  
A/Reference number: Z16552  
A/Accession: T09356  
A/Molecule type: DNA  
A/Residues: 1-1196 <REV>  
A/Cross-References: EMBL:AU078620; GSPDB:GNO0062; ATSP:F23K16.30  
A/Experimental source: cultivar Columbia; BAC clone F23K16  
C/Genetics:  
A/Gene: ATSP:F23K16.30; BRI1  
A/Map position: 4  
C/Superfamily: protein kinase Xa21; leucine-rich alpha-2-glycoprotein repeat homology; p

Query Match 100.0%; Score 332; DB 2; Length 1196;  
Best Local Similarity 100.0%; Pred. No. 5.4e-34;  
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 AANFIAGKRVYIKNDGKMKCHGAGNLEFGQIRSHQNLSTRNPNITSRYVGGHTS 60  
588 AANFIAGKRVYIKNDGKMKCHGAGNLEFGQIRSHQNLSTRNPNITSRYVGGHTS 647

Oy 61 PT 62  
Db 648 PT 649

RESULT 2  
P06598  
protein F20N2.4 [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C/Accession: P06598  
R/Theologis, A.; Ecker, J. R.; Palm, C. J.; Federspiel, N. A.; Kaul, S.; White, O.; Alonso, C.; Chiu, C. W.; Chung, M. K.; Conn, L.; Conway, A. B.; Conway, A. R.; Creasy, T. H.; Dewar, K.; Jensen, N. F.; Hughes, B.; Hulzar, L.  
Nature 408, 816-820, 2000  
A/Authors: Hunter, J. L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C. C.; Li, J. H.; Li, Y.; Lin, X.; Liu, S. X.; Liu, Z. A.; Luvos, J. S.; Maiti, R.; Marziani, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A/Authors: Salzberg, S. L.; Schwartz, J. R.; Shinn, P.; Southwick, A. M.; Sun, H.; Tallon, I. Ker, M.; Wu, D.; Yu, G.; Fraser, C. M.; Venter, J. C.; Davis, R. W.  
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A/Reference number: A61141; MUID:21016719; PMID:11130712  
A/Accession: P06598  
A/Status: preliminary

#### ALIGNMENTS

30	55.5	16.7	658	2	G98101	transketolase (EC
31	55.5	16.7	1100	2	H71616	SERA antigen/papai
32	55.5	16.7	1192	2	F88062	protein ZK250.10 [
33	55.5	16.7	1360	2	T32833	hypothetical prote
34	55.5	16.7	1466	2	T32422	hypothetical prote
35	55	16.6	322	2	G89926	hypothetical prote
36	55	16.6	359	2	G95955	conserved probable
37	55	16.6	445	2	T20190	hypothetical prote
38	55	16.6	467	2	A57627	p55 erythrocyte me
39	55	16.6	474	2	E64207	hypothetical prote
40	55	16.6	890	2	A48753	hypothetical prote
41	54.5	16.4	384	2	F32252	gene III protein -
42	54.5	16.4	414	2	T23940	hypothetical prote
43	54.5	16.4	573	2	A41319	3-oxosteroid 1-deh
44	54	16.3	155	2	C97052	choistmate mutase
45	54	16.3	211	2	B84066	hypothetical prote



M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, DNA Res. 5, 55-76, 1998  
 A>Title: Complete sequence and gene organization of the genome of a hyper-thermophilic  
 A/Reference number: A71000; MUID:98344137; PMID:9679194  
 A/Accession: B71039  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-871 <K&M>  
 A/Cross-references: GB:AP00006; NID:93336133; PIDN:BA30714.1; PID:93258031  
 A/Experimental source: strain OT3  
 A/Note: this accession replaces an interim accession for a sequence replaced by GenBank  
 C/Genetics:  
 A/Gene: PH1602

Query Match 18.7%; Score 62; DB 2; Length 871;  
 Best Local Similarity 37.0%; Pred. No. 14;  
 Matches 17; Conservative 6; Mismatches 9; Indels 14; Gaps 3;  
 Oy 10 YVYIKNDGKKE-----CHGAGNLT-----EFG--IRSEQLNR 41  
 D 776 YILAGTGAKEITFGSCGAGNVLKRAKATRYRDRIRQLNLNR 821

RESULT 8  
 A64084  
 Iysohospholipase I2 homolog - Haemophilus influenzae (strain Rd KW20)  
 C/Species: Haemophilus influenzae  
 C/Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 08-Oct-1999  
 C/Accession: A64084  
 R/Plaschmann, R.D.; Adams, M.D.; White, O.; Kihres, E.F.; Kerlavage, J.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodex, A.; Kelley, J.M.; Weidman, J.D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M. Science 269, 496-512, 1995  
 A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A./Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
 A/Reference number: A64080; MUID:95350630; PMID:7542800  
 A/Accession: A64084  
 A/Status: nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-313 <TIGR>  
 A/Cross-references: GB:U32747; GB:I42023; NID:91573635; PIDN:AC22305.1; PID:91573643; T C/Superfamily: Iysohospholipase I2

Query Match 18.4%; Score 61; DB 2; Length 313;  
 Best Local Similarity 36.6%; Pred. No. 6;  
 Matches 15; Conservative 9; Mismatches 15; Indels 2; Gaps 1;  
 Oy 7 GRRYVYIKNDGKKECHGAGNLTLEFGIRSEQLNRSTRNP 47  
 D 181 GRRYVFGK--GAYQQAHLVYVELTFCRTRKMRNRINRKNP 219

RESULT 9  
 AE1810  
 WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)  
 C/Species: Nostoc sp.  
 A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C/Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
 C/Accession: AE1810  
 R/Kanezo, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 9, 205-213, 2001  
 A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
 A/Reference number: AB1807; MUID:21595285; PMID:11759840  
 A/Accession: AE1810  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-1227 <KUR>  
 A/Cross-references: GB:BA000019; PIDN:BA077553.1; PID:917135007; GSPDB:GN00179  
 A/Experimental source: strain PCC 7120  
 C/Genetics:  
 A/Gene: alr0029

Query Match 18.1%; Score 60; DB 2; Length 1227;  
 Best Local Similarity 25.5%; Pred. No. 36;  
 Matches 12; Conservative 16; Mismatches 17; Indels 2; Gaps 1;  
 Oy 13 IKNDGKKECHGAGNLTLEFGIRSEQLNRSTRNPNTSRVYGH 59  
 Db 528 LQKEPLPKGYAKGNLNL--LRQLQDRIPEPSFIDSGRDFSLT 572

RESULT 10  
 T28432  
 variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)  
 N/Alternate names: erythrocyte membrane binding protein 1 (EMPI)  
 C/Species: Plasmodium falciparum  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000  
 C/Accession: T28432  
 R/Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herfeldt, J.A.; Peterson, D.S. Cell 82, 89-100, 1995  
 A/Title: The large diverse gene family var encodes proteins involved in cytoadherence and  
 A/Reference number: Z20487; MUID:95330813; PMID:7606788  
 A/Accession: T28432  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-3078 <STX>  
 A/Cross-references: EMBL:L40608; NID:9886374; PID:9886375; PIDN:AAA75396.1  
 C/Genetics:  
 A/Gene: var-1  
 A/Introns: 2611/3

Query Match 18.1%; Score 60; DB 2; Length 3078;  
 Best Local Similarity 31.4%; Pred. No. 99;  
 Matches 16; Conservative 6; Mismatches 15; Indels 14; Gaps 3;  
 Oy 16 DQMKKECHGAGNLTLEFGIRSEQLNRSTRNPNC-----NITSRYVGH 58  
 Db 2278 DQLENACRGAG---IFEGIRDE--WKCRNVCGYVCKPENVNGEAKGH 2322

RESULT 11  
 H66716  
 L-asparaginase [imported] - Lactococcus lactis subsp. lactis (strain IL1403)  
 C/Species: Lactococcus lactis subsp. lactis  
 C/Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
 C/Accession: H66716  
 R/Bolotin, A.; Winkler, P.; Mauger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehrlich, Genome Res. 11, 731-753, 2001  
 A/Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp  
 A/Reference number: A86625; MUID:21235186; PMID:11337471  
 A/Accession: H66716  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-323 <STO>  
 A/Cross-references: GB:AE005176; PID:912723651; PIDN:AAK04834.1; GSPDB:GN00146  
 A/Experimental source: strain IL1403  
 C/Genetics:  
 A/Gene: amB  
 C/Superfamily: asparaginase

Query Match 17.8%; Score 59; DB 2; Length 323;  
 Best Local Similarity 38.3%; Pred. No. 11;  
 Matches 18; Conservative 5; Mismatches 20; Indels 4; Gaps 2;  
 Oy 16 DQMKKECHGAGNLTLEFGIRSEQLNR-LSTRNPNTSRVYGH 61  
 Db 229 DGVIVLALGAGN---PLASQSLQRLDAPVVLVSRCFGIAEP 272

RESULT 12  
 S15799  
 protein-tyrosine-phosphatase (EC 3.1.3.48) cdc25 homolog - Caenorhabditis elegans  
 N/Alternate names: protein ZK637.11  
 C/Species: Caenorhabditis elegans  
 C/Date: 06-Jan-1995 #sequence\_revision 25-Apr-1997 #text\_change 20-Mar-1998

C/Accession: S15799  
 R/Citation: M.; Almscough, R.; Coulson, A.; Dear, S.; Du, Z.; Durbin, R.; Green, P.; Hall, R.; Staden, R.; Sulston, J.; Thierry-Mieg, J.; Thomas, K.; Waterston, R.; Wilson, R.  
 submitted to the EMBL Data Library, May 1991  
 A/Reference number: S15786  
 A/Accession: S15799  
 A/Molecule type: DNA  
 A/Residues: 1-316 <GRA>  
 A/Cross-references: EMBL:Z11115, NID:96953, PID:96964  
 C/Genetics:  
 A/Introns: 34/2; 157/3  
 C/Function:  
 A/Description: catalyses hydrolysis of peptidyl-phosphotyrosine to peptidyl-tyrosine and A/Pathway: initiation of mitosis  
 A/Note: cdc25 activates the cdc2 protein kinase by dephosphorylating it  
 C/Superfamily: Caenothadit protein-tyrosine-phosphatase cdc25 homolog; cdc25-type pro C/Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatase F/177-274/Domain: cdc25-type protein-tyrosine-phosphatase homology <PRP>  
 F/189/Active site: Cys (phosphocysteine intermediate) #status predicted  
 F/Binding site: substrate phosphate (Arg) #status predicted

Query Match 17.6%; Score 58.5; DB 1; Length 316;  
 Best Local Similarity 27.3%; Pred. No. 13;  
 Matches 15; Conservative 13; Mismatches 22; Indels 5; Gaps 2;

QY 13 IKNDGKKEC---HGAGNLEFGIR-SEQLNRLSTRNPNCTSRVYGHTSPT 62  
 DB 12 VRNDGLRLKSECEAGSSKLPFRQNRHSSAIIHINSSPPTKRSDIGGYTGT 66

RESULT 13  
 C69072  
 A/Description: ribonucleoside-triphosphate reductase - Methanobacterium thermoautotrophicum C/Species: Methanobacterium thermoautotrophicum  
 C/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
 C/Accession: C69072  
 R/Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; R/Suth, D.; Spadafora, R.; Vicaltre, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
 J. Bacteriol. 179, 7135-7155, 1997  
 A/Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct A/Reference number: A69000; MUID:9803514; PMID:93711463  
 A/Accession: C69072  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-776 <MTH>  
 A/Cross-references: GB:AE000914; GB:AE000666; NID:92622656; PIDN:AAB86013.1; PID:9262265  
 A/Experimental source: strain Delta H  
 A/Genetics:  
 A/Name: MTH1539  
 A/Start codon: TTG

Query Match 17.6%; Score 58.5; DB 2; Length 776;  
 Best Local Similarity 36.4%; Pred. No. 34;  
 Matches 16; Conservative 6; Mismatches 19; Indels 3; Gaps 1;

QY 5 IAGKRYVIKNDGKKECHGAGNLEFG---GIRSEQLNRLSTR 45  
 DB 541 IDGERYRIENATMSPFTGLNMLEYHLAGIGSPENRFGRL 584

RESULT 14  
 S23764  
 A/Description: polygalacturonase-inhibiting protein precursor - kidney bean  
 C/Species: Phaseolus vulgaris (kidney bean)  
 C/Date: 05-Mar-1994 #sequence\_revision 10-Nov-1995 #text\_change 21-Jan-2000  
 C/Accession: S23764  
 R/Toubart, P.; Desiderio, A.; Salvi, G.; Cervone, F.; Daroda, L.; de Lorenzo, G.; Bergma Plant J. 2, 367-373, 1992  
 A/Title: Cloning and characterization of the gene encoding the endopolygalacturonase-int A/Reference number: S23764; MUID:93272053; PMID:1303801  
 A/Accession: S23764  
 A/Status: preliminary

A/Molecule type: DNA  
 A/Residues: 1-342 <TOU>  
 A/Cross-references: EMBL:X64769, NID:921028; PIDN:CAA6016.1; PID:921029  
 A/Note: It is uncertain whether Met-1 or Met-10 is the initiator  
 C/Superfamily: polygalacturonase-inhibiting protein; leucine-rich alpha-2-glycoprotein re F/274-297/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR>

Query Match 17.5%; Score 58; DB 2; Length 342;  
 Best Local Similarity 34.5%; Pred. No. 16;  
 Matches 20; Conservative 7; Mismatches 21; Indels 10; Gaps 3;

QY 3 NFIAKRYVIKNDGKKECHGAGNLEFG---GIRSEQLNRLSTRNPNCTSRVYG 56  
 DB 236 NMEBGASVIFGSDKTKTKTHLAKSLAFDLGVGH-SKLNGLDIRN-----NRIYG 287

RESULT 15  
 JE0082  
 A/Description: G-protein coupled receptor precursor - mouse  
 N/Alternate names: GFR1alpha-3  
 C/Species: Mus musculus (house mouse)  
 C/Date: 21-May-1998 #sequence\_revision 29-May-1998 #text\_change 02-Jun-2000  
 C/Accession: JE0082  
 R/Nomoto, S.; Ito, S.; Yang, L.X.; Kuchi, K.  
 Biochem. Biophys. Res. Commun. 244, 849-853, 1998  
 A/Title: Molecular cloning and expression analysis of GFR1alpha-3, a novel cDNA related to A/Reference number: JE0082; MUID:98205811; PMID:9535755  
 A/Accession: JE0082  
 A/Molecule type: mRNA  
 A/Residues: 1-397 <NOM>  
 A/Cross-references: DDBJ:AB008833; NID:92627159; PIDN:BA23562.1; PID:92627160  
 A/Comment: This protein plays a distinct role in cell survival and differentiation.  
 C/Superfamily: Mus musculus G-protein coupled receptor  
 C/Keywords: glycoprotein  
 F/1-25/Domain: signal sequence #status predicted <SIG>  
 F/380-397/Region: hydrophobic  
 F/92,145,306/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 17.5%; Score 58; DB 2; Length 397;  
 Best Local Similarity 44.4%; Pred. No. 19;  
 Matches 12; Conservative 4; Mismatches 7; Indels 4; Gaps 1;

QY 22 CHGAGNLEFGIRSEQLNRLSTRNPC 48  
 DB 315 CRGAGNLD---ECQJERSFSPQPC 337

Search completed: March 10, 2003, 18:24:54  
 Job time: 3.77424 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 17:51:34 ; Search time 1.33068 Seconds

(without alignments)  
1964.817 Million cell updates/sec

Title: US-09-823-394-2\_COPY\_588\_649

Sequence: 1 AANFIAGKRRVYIKNDGMRK.....STRNPNITSRVYGHTSPT 62

Scoring table: BLOSUM62

Gapop 10.0 , Gapept 0.5

Searched: 188354 seqs, 42170167 residues

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: Published Applications AA:  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	332	100.0	1196 9 US-09-823-394-2	Sequence 2, Appl1
2	60.5	18.2	466 10 US-09-825-414-20	Sequence 20, Appl1
3	60.5	18.2	808 9 US-10-108-605-127	Sequence 127, Appl1
4	60	18.1	2710 9 US-10-153-273-12	Sequence 12, Appl1
5	58	17.5	397 10 US-09-220-920-64	Sequence 64, Appl1
6	57.5	17.3	732 9 US-09-991-496-18	Sequence 18, Appl1
7	57.5	17.3	732 10 US-09-932-257A-25	Sequence 25, Appl1
8	57.5	17.3	732 10 US-09-759-010-5	Sequence 5, Appl1
9	57.5	17.3	732 10 US-09-874-923-18	Sequence 18, Appl1
10	57	17.2	386 9 US-10-007-262-1	Sequence 1, Appl1
11	57	17.2	386 10 US-09-816-825-2	Sequence 2, Appl1
12	54.5	16.4	230 10 US-09-988-982-3	Sequence 3, Appl1
13	52	15.7	62 9 US-09-796-692-1724	Sequence 1724, Appl1
14	52	15.7	63 10 US-09-864-761-48451	Sequence 48451, Appl1
15	52	15.7	229 12 US-10-038-107A-3	Sequence 891, Appl1
16	52	15.7	257 9 US-09-925-299-891	Sequence 891, Appl1
17	52	15.7	257 10 US-09-925-299-891	Sequence 891, Appl1
18	51.5	15.5	208 10 US-09-988-982-1	Sequence 1, Appl1
19	51.5	15.5	263 9 US-09-925-299-991	Sequence 991, Appl1

20	51.5	15.5	263 10 US-09-925-299-991	Sequence 991, Appl1
21	51.5	15.5	331 9 US-10-166-580-2	Sequence 2, Appl1
22	51	15.4	123 10 US-09-764-898-242	Sequence 242, Appl1
23	51	15.4	294 10 US-09-815-242-13076	Sequence 13076, Appl1
24	51	15.4	340 10 US-09-764-898-169	Sequence 169, Appl1
25	51	15.4	379 9 US-10-012-896-946	Sequence 946, Appl1
26	51	15.4	379 9 US-09-895-814-946	Sequence 946, Appl1
27	51	15.4	518 9 US-10-098-174-1	Sequence 1, Appl1
28	51	15.4	690 10 US-09-815-242-5841	Sequence 5841, Appl1
29	51	15.4	1185 9 US-09-895-913A-246	Sequence 246, Appl1
30	51	15.4	1724 9 US-09-964-899-43	Sequence 43, Appl1
31	51	15.4	2150 10 US-09-321-987B-2	Sequence 2, Appl1
32	51	15.4	2165 10 US-09-800-729-155	Sequence 155, Appl1
33	50.5	15.2	67 10 US-09-925-301-968	Sequence 968, Appl1
34	50.5	15.2	358 10 US-09-823-356-6	Sequence 6, Appl1
35	50.5	15.2	358 10 US-09-740-027-4	Sequence 4, Appl1
36	50.5	15.2	434 12 US-10-001-843-179	Sequence 179, Appl1
37	50.5	15.2	491 9 US-09-895-913A-106	Sequence 106, Appl1
38	50.5	15.2	647 9 US-09-991-262-50	Sequence 50, Appl1
39	50.5	15.2	675 9 US-09-991-262-52	Sequence 52, Appl1
40	50.5	15.2	1156 10 US-09-757-716-1	Sequence 1, Appl1
41	50	15.1	400 10 US-09-220-920-63	Sequence 63, Appl1
42	50	15.1	400 10 US-09-828-366-16	Sequence 16, Appl1
43	50	15.1	617 9 US-10-012-896-947	Sequence 947, Appl1
44	50	15.1	617 9 US-09-895-814-947	Sequence 947, Appl1
45	50	15.1	914 10 US-09-815-242-4996	Sequence 4996, Appl1

#### ALIGNMENTS

RESULT 1  
US-09-823-394-2 Application US/09823394  
Publication No. US20030041344A1  
GENERAL INFORMATION:  
APPLICANT: Chory, Joanne  
APPLICANT: Jilmaning, Li  
TITLE OF INVENTION: Salk Institute for Biological Studies  
FILE REFERENCE: SALKINS.012CPI  
CURRENT APPLICATION NUMBER: US/09/823.394  
CURRENT FILING DATE: 2001-03-30  
PRIOR APPLICATION NUMBER: 08/881,706  
PRIOR FILING DATE: 1997-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 1196  
TYPE: PRT  
ORGANISM: Arabidopsis  
US-09-823-394-2

Query Match 100.0%; Score 332; DB 9; Length 1196;  
Best Local Similarity 100.0%; Pred. No. 7.3e-36;  
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AANFIAGKRRVYIKNDGMRKCHAGNLEFQGRSOLNRLSTRNPNITSRVYGHTS 60  
DB 588 AANFIAGKRRVYIKNDGMRKCHAGNLEFQGRSOLNRLSTRNPNITSRVYGHTS 647

QY 61 PT 62  
DB 648 PT 649

RESULT 2  
US-09-825-414-20  
Sequence 20, Application US/09825414  
Patent No. US20020083489A1  
GENERAL INFORMATION:  
APPLICANT: Collier, Alan  
APPLICANT: Alfano, James R.

APPLICANT: Charkowski, Amy O.  
TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE  
TITLE OR INVENTION: HRP PATHOGENICITY ISLAND AND THEIR USES  
FILE REFERENCE: 19603/3243  
CURRENT APPLICATION NUMBER: US/09/825,414  
PRIOR FILING DATE: 2001-04-03  
PRIOR APPLICATION NUMBER: 60/194,160  
PRIOR FILING DATE: 2000-04-03  
PRIOR APPLICATION NUMBER: 60/224,604  
PRIOR FILING DATE: 2000-08-11  
PRIOR APPLICATION NUMBER: 60/249,548  
PRIOR FILING DATE: 2000-11-17  
NUMBER OF SEQ ID NOS: 91  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 20  
LENGTH: 466  
TYPE: PRT  
ORGANISM: Pseudomonas syringae  
US-09-825-414-20

Query Match  
Best Local Similarity 18.2%; Score 60.5; DB 10; Length 466;  
St Local Similarity 26.4%; Pred. No. 4.3;  
Matches 19; Conservative 11; Mismatches 31; Indels 11; Gaps 2;

Qy 1 AANFAGKRYVYIKNDG-----KKECHGAGNLEFGGIRSEQLN--RLSTRNPN 49  
Db 387 AMNYVAAKIRLSKPEGKVAFAVGATHTATSCDVGPLAELHGVSLVIDLGLKSRATVD 446

Qy 50 ITRVYGGHTSP 61  
Db 447 INVKNYGGKLNK 458

RESULT 3  
US-10-108-605-127  
Sequence 127; Application US/10108605  
Patent No. US20020160934A1  
GENERAL INFORMATION:  
APPLICANT: Broadus, Julie  
APPLICANT: Stam, Lynn  
APPLICANT: Bachmann, Jane  
APPLICANT: Kamdar, Kim  
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE  
TITLE OF INVENTION: PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF  
FILE REFERENCE: 31133B  
CURRENT APPLICATION NUMBER: US/10/108,605  
CURRENT FILING DATE: 2002-03-27  
PRIOR APPLICATION NUMBER: US 09/761,142  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/176,418  
PRIOR FILING DATE: 2000-01-14  
NUMBER OF SEQ ID NOS: 361  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 127  
LENGTH: 808  
TYPE: PRT  
ORGANISM: Drosophila melanogaster  
US-10-108-605-127

Query Match  
Best Local Similarity 18.2%; Score 60.5; DB 9; Length 808;  
St Local Similarity 38.0%; Pred. No. 8.4;  
Matches 19; Conservative 5; Mismatches 23; Indels 3; Gaps 1;

Qy 2 ANFIAGKRYVYIKNDGKKECHGAGNLEFGGIRSEQLNRLSTRNPN 51  
Db 727 ARFKQGRVYVYEDNGINP---GAFNPLQOIRKRYTLARLLCDNSDRLL 773

RESULT 4  
US-10-153-273-12  
Sequence 12; Application US/10153273  
Patent No. US20020169305A1  
GENERAL INFORMATION:

APPLICANT: Sim, Kim L.  
Chitnis, Chetan  
Miller, Louis H.  
Peterson, David S.  
Su, Xin-zhaun  
Wellems, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/153,273  
FILING DATE: 21-May-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/210,288  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Fuller, Michael  
REGISTRATION NUMBER: 36,516  
REFERENCE/DOCKET NUMBER: NIH121.15WDV1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2710 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium falciparum  
SEQUENCE DESCRIPTION: SEQ ID NO: 12:  
US-10-153-273-12

Query Match  
Best Local Similarity 18.1%; Score 60; DB 9; Length 2710;  
St Local Similarity 31.4%; Pred. No. 43;  
Matches 16; Conservative 6; Mismatches 15; Indels 14; Gaps 3;

Qy 16 DAKKKECHGAGNLEFGGIRSEQLNRLSTRNPN-----NITSRVYGGH 58  
Db 2280 DGLNNACRGAG---IFGIRKDE---WKCRNVCGYVCKPEVNVNGAKKH 2324

RESULT 5  
US-09-220-920-64  
Sequence 64; Application US/09220920  
Patent No. US2002002269A1  
GENERAL INFORMATION:  
APPLICANT: Milbrandt, Jeffrey D.  
APPLICANT: Balogh, Robert H.  
TITLE OF INVENTION: Artemin, A No. US2002002269A1 Neurotrophic Factor  
FILE REFERENCE: 6029-7996  
CURRENT APPLICATION NUMBER: US/09/220,920  
CURRENT FILING DATE: 1998-12-24  
EARLIER APPLICATION NUMBER: 09/163,283  
EARLIER FILING DATE: 1998-09-29  
EARLIER APPLICATION NUMBER: 60/108,148  
EARLIER FILING DATE: 1998-11-12  
EARLIER APPLICATION NUMBER: 09/218,698

EARLIER FILING DATE: 1998-12-22  
NUMBER OF SEQ ID NOS: 120  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 64  
LENGTH: 397  
TYPE: PRT  
ORGANISM: Murine  
US-09-220-920-64

Query Match 17.5%; Score 58; DB 10; Length 397;  
Best Local Similarity 44.4%; Pred. No. 7.6;  
Matches 12; Conservative 4; Mismatches 7; Indels 4; Gaps 1;

QY 22 CHGAGNLEFGIRSEQLNR--LSTRNPNITSRVYG 48  
DB 315 CRGSGNLD---KCKIKKYIDGELNKTPIWTRNPDITNEBYG 317

RESULT 6  
US-09-496-18  
Sequence 18, Application US/09991496  
Patent No. US20020169285A1  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Campos-Melo, Antonio  
APPLICANT: Webb, John R.  
APPLICANT: Dillon, David C.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Bhatia, Ajay  
APPLICANT: Coler, Rhea  
APPLICANT: Probst, Peter  
APPLICANT: Brannon, Mark  
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
FILE REFERENCE: 210121.420C9  
CURRENT APPLICATION NUMBER: US/09/991,496  
CURRENT FILING DATE: 2001-11-20  
NUMBER OF SEQ ID NOS: 137  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 18  
LENGTH: 732  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-991-496-18

Query Match 17.3%; Score 57.5; DB 9; Length 732;  
Best Local Similarity 37.0%; Pred. No. 19;  
Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;  
QY 14 KNDGKKECHGAGNLEFGIRSEQLNR--LSTRNPNITSRVYG 56  
DB 269 KDDGDKK---KKIKKYIDGELNKTPIWTRNPDITNEBYG 310

RESULT 7  
US-09-932-257A-25  
Sequence 25, Application US/09932257A  
Publication No. US20030039658A1  
GENERAL INFORMATION:  
APPLICANT: Estable, Mario  
APPLICANT: Koeder, Robert  
TITLE OF INVENTION: MCEF, A No. US20030039658A1 Transcription Factor  
FILE REFERENCE: 600-1-269N  
CURRENT APPLICATION NUMBER: US/09/932,257A  
CURRENT FILING DATE: 2001-08-17  
PRIOR APPLICATION NUMBER: 60/226,340  
PRIOR FILING DATE: 2000-08-18  
PRIOR APPLICATION NUMBER: 60/226,339  
PRIOR FILING DATE: 2000-08-18  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 25  
LENGTH: 732

TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-932-257A-25

Query Match 17.3%; Score 57.5; DB 9; Length 732;  
Best Local Similarity 37.0%; Pred. No. 19;  
Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;

QY 14 KNDGKKECHGAGNLEFGIRSEQLNR--LSTRNPNITSRVYG 56  
DB 269 KDDGDKK---KKIKKYIDGELNKTPIWTRNPDITNEBYG 310

RESULT 8  
US-09-759-010-5  
Sequence 5, Application US/09759010  
Patent No. US20010034042A1  
GENERAL INFORMATION:  
APPLICANT: Srivastava, Pramod K.  
TITLE OF INVENTION: COMPLEXES OF PEPTIDE BINDING FRAGMENTS OF HEAT-SHOCK  
FILE REFERENCE: 8449-135  
CURRENT APPLICATION NUMBER: US/09/759,010  
CURRENT FILING DATE: 2001-01-12  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 732  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-759-010-5

Query Match 17.3%; Score 57.5; DB 10; Length 732;  
Best Local Similarity 37.0%; Pred. No. 19;  
Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;  
QY 14 KNDGKKECHGAGNLEFGIRSEQLNR--LSTRNPNITSRVYG 56  
DB 269 KDDGDKK---KKIKKYIDGELNKTPIWTRNPDITNEBYG 310

RESULT 9  
US-09-874-923-18  
Sequence 18, Application US/09874923  
Patent No. US20020081320A1  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Campos-Melo, Antonio  
APPLICANT: Webb, John R.  
APPLICANT: Dillon, David C.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Bhatia, Ajay  
APPLICANT: Coler, Rhea  
APPLICANT: Probst, Peter  
APPLICANT: Brannon, Mark  
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
FILE REFERENCE: 210121.420C8  
CURRENT APPLICATION NUMBER: US/09/874,923  
CURRENT FILING DATE: 2001-06-04  
NUMBER OF SEQ ID NOS: 122  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 18  
LENGTH: 732  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-874-923-18

Query Match 17.3%; Score 57.5; DB 10; Length 732;  
Best Local Similarity 37.0%; Pred. No. 19;  
Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;  
QY 14 KNDGKKECHGAGNLEFGIRSEQLNR--LSTRNPNITSRVYG 56

Db 269 KKGDKKK-----KKKIKKYYIDKEELNKTKPIWTRNPDITNEYG 310

RESULT 10  
US-10-007-262-1  
Sequence 1, Application US/10007262

Patent No. US20020164748A1

GENERAL INFORMATION:

APPLICANT: Bistrup, Annette

APPLICANT: Rosen, Steven D.

APPLICANT: Tangemann, Kirsten

APPLICANT: Hemmerich, Stefan

TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3

FILE REFERENCE: 6510-107CIP

CURRENT APPLICATION NUMBER: US/10/007,262

CURRENT FILING DATE: 2001-11-08

PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911

PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12

NUMBER OF SEQ ID NOS: 8

SOFTWARE: FastSeq for Windows Version 3.0

ID NO 1

LENGTH: 386

TYPE: PRT

ORGANISM: H. sapiens

US-10-007-262-1

Query Match

Best Local Similarity 41.0%; Pred. No. 9.9;

Matches 16; Conservative 6; Mismatches 11; Indels 6; Gaps 1;

Y 12 YIKNDGKKECHGAGNLLFQGISRSEQ-----LNRLST 44

Db 340 YEKVSRLOKACGDAMNLLGYRHVRSSEQQRNLLDLST 378

RESULT 11  
US-09-816-825-2  
Sequence 2, Application US/09816825

Patent No. US20010051370A1

GENERAL INFORMATION:

APPLICANT: Bistrup, Annette

APPLICANT: Rosen, Steven D.

APPLICANT: Hemmerich, Stefan

TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3

FILE REFERENCE: 6510-107CON

CURRENT APPLICATION NUMBER: US/09/816,825

CURRENT FILING DATE: 2001-03-22

PRIOR APPLICATION NUMBER: 09/045,284

PRIOR FILING DATE: 1998-03-20

NUMBER OF SEQ ID NOS: 9

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 2

LENGTH: 386

TYPE: PRT

ORGANISM: Homo sapiens

US-09-816-825-2

Query Match

Best Local Similarity 41.0%; Pred. No. 9.9;

Matches 16; Conservative 6; Mismatches 11; Indels 6; Gaps 1;

Y 12 YIKNDGKKECHGAGNLLFQGISRSEQ-----LNRLST 44

Db 340 YEKVSRLOKACGDAMNLLGYRHVRSSEQQRNLLDLST 378

RESULT 12  
US-09-988-982-3  
Sequence 3, Application US/09988982

Patent No. US20020081699A1

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

Shah, Puri  
Murry, Lynn E.

TITLE OF INVENTION: NOVEL HUMAN LYSOPHOSPHOLIPASE

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/988,982

FILING DATE: 19-NO. US20020081699A1-2001

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/213,394

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0269 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-845-4166

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 230 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: GenBank

CLONE: 552244

SEQUENCE DESCRIPTION: SEQ ID NO: 3:

US-09-988-982-3

Query Match

Best Local Similarity 16.4%; Score 54.5; DB 10; Length 230;

Matches 17; Conservative 7; Mismatches 14; Indels 7; Gaps 3;

Y 21 ECHGAGN---LLEFGIRSEQLNRLSTRPCNITSRVYGG--HTS 60

Db 168 QCHGDCDPLVPLMFGSLTYERLKGIV--NPANVTREYEGMMHSS 210

RESULT 13  
US-09-796-692-1724

Sequence 1724, Application US/09796692

Publication No. US20020198362A1

GENERAL INFORMATION:

APPLICANT: Gaiger, Alexander

APPLICANT: Algate, Paul A.

APPLICANT: Mammon, Jane

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY

FILE REFERENCE: 2077.001200

CURRENT APPLICATION NUMBER: US/09/796,692

CURRENT FILING DATE: 2001-03-01

PRIOR APPLICATION NUMBER: 60/186,126

PRIOR FILING DATE: 2000-03-01

PRIOR APPLICATION NUMBER: 60/190,479

PRIOR FILING DATE: 2000-03-17

PRIOR APPLICATION NUMBER: 60/200,545

PRIOR FILING DATE: 2000-04-27

PRIOR APPLICATION NUMBER: 60/200,303

PRIOR FILING DATE: 2000-04-28

PRIOR APPLICATION NUMBER: 60/200,779

PRIOR FILING DATE: 2000-04-28



PRIOR APPLICATION NUMBER: 60/200,999  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/202,084  
PRIOR FILING DATE: 2000-05-04  
PRIOR APPLICATION NUMBER: 60/206,201  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: 60/218,950  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: 60/222,903  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: 60/223,416  
PRIOR FILING DATE: 2000-08-04  
PRIOR APPLICATION NUMBER: 60/223,378  
PRIOR FILING DATE: 2000-08-07  
NUMBER OF SEQ ID NOS: 9597  
SOFTWARE: FaastSeq for Windows Version 3.0  
SEQ ID NO 1724  
LENGTH: 62  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-864-761-48451

Query Match 15.7%; Score 52; DB 9; Length 62;  
Best Local Similarity 44.8%; Pred. No. 4.9;  
Matches 13; Conservative 1; Mismatches 15; Indels 0; Gaps 0;

29 LEEFGIRSEQLNLTSTNPNITSRVYGG 57  
19 LTTGIRFEDNGIYFCQCKNNTSEVYGG 47

## RESULT 14

US-09-864-761-48451  
Sequence 48451, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecm1ca-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
SEQ ID NO 48451  
LENGTH: 63  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC004859.2  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.77  
OTHER INFORMATION: EST HUMAN HIT: AA077526.1, EVALU 8.00e-09  
OTHER INFORMATION: SWISSPROT HIT: Q13751, EVALU 4.40e+00  
US-09-864-761-48451

Query Match 15.7%; Score 52; DB 10; Length 63;  
Best Local Similarity 31.2%; Pred. No. 5;  
Matches 10; Conservative 8; Mismatches 12; Indels 2; Gaps 1;

10 YYIINDGKKKCH--GAGNLEFGIRSEQL 39  
19 YSLVGCGAHHRQCHGVGVLDPTGLVSRV 50

## RESULT 15

US-10-038-107A-3  
Sequence 3, Application US/10038107A  
Patent No. US20020150573A1  
GENERAL INFORMATION:  
APPLICANT: Nusse, Michael  
TITLE OF INVENTION: ANTI-19 ALPHA-BETA ANTIBODY FOR LYMPHOMA THERAPY  
FILE REFERENCE: 7529/0H405  
CURRENT APPLICATION NUMBER: US/10/038,107A  
CURRENT FILING DATE: 2002-04-17  
PRIOR APPLICATION NUMBER: 60/247,079  
PRIOR FILING DATE: 2000-10-11  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3  
LENGTH: 229  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-038-107A-3

Query Match 15.7%; Score 52; DB 12; Length 229;  
Best Local Similarity 44.8%; Pred. No. 24;  
Matches 13; Conservative 1; Mismatches 15; Indels 0; Gaps 0;

29 LEEFGIRSEQLNLTSTNPNITSRVYGG 57  
107 LTTGIRFEDNGIYFCQCKNNTSEVYGG 135

Search completed: March 10, 2003, 18:11:43  
Job time: 2.33068 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 18:03:39 ; Search time 1.33068 Seconds  
(Without alignments)  
1370.890 Million cell updates/sec

Title: US-09-823-394-2\_COPY\_588\_649  
Sequence: 332  
Perfect score: 1 AANFIAGKRYVYIKNDGKMK.....STRNPNITSRVYGHTSPT 62

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: Issued Patents AA:\*  
2: /cgn2\_6/prodata/2/1aa/5A COMB pep:\*  
3: /cgn2\_6/prodata/2/1aa/5B COMB pep:\*  
4: /cgn2\_6/prodata/2/1aa/6A COMB pep:\*  
5: /cgn2\_6/prodata/2/1aa/6B COMB pep:\*  
6: /cgn2\_6/prodata/2/1aa/PCTUS COMB pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	332	100.0	1196	4	US-08-881-706-2
2	60	18.1	2710	2	US-08-568-459A-12
3	60	18.1	2710	2	US-08-487-826B-12
4	60	18.1	2710	4	US-09-210-288-12
5	60	18.1	3060	2	US-08-487-826B-14
6	58	17.5	342	1	US-08-244-646-15
7	58	17.5	342	1	US-08-592-936B-21
8	58	17.5	342	2	US-09-111-573-21
9	58	17.5	397	4	US-09-220-528-64
10	57.5	17.3	732	2	US-08-533-669A-18
11	57.5	17.3	732	4	US-09-307-143-4
12	57.5	17.3	732	4	US-09-183-861-18
13	57.5	17.3	732	4	US-09-022-765-18
14	57.5	17.3	732	4	US-09-045-284A-2
15	57	17.2	386	4	US-09-190-911-1
16	55	16.6	130	2	US-08-467-046-12
17	55	16.6	890	1	US-08-145-006C-5
18	55	16.6	890	5	PCT-US94-00545-5
19	54.5	16.4	230	2	US-08-844-120-3
20	54.5	16.4	230	2	US-09-022-940-5
21	54.5	16.4	230	3	US-09-216-001-3
22	54.5	16.4	230	3	US-09-216-386-5
23	54.5	16.4	230	4	US-08-878-862-3
24	54.5	16.4	230	4	US-09-213-394-3
25	54.5	16.4	573	1	US-08-200-512-2
26	54	16.3	227	1	US-08-244-646-17
27	53	16.0	432	2	US-08-700-152A-4

28	53	16.0	525	1	US-08-077-939-19	Sequence 19, Appl
29	53	16.0	525	1	US-08-461-599-19	Sequence 19, Appl
30	53	16.0	525	1	US-08-461-621-19	Sequence 19, Appl
31	53	16.0	525	1	US-08-465-334-19	Sequence 19, Appl
32	52	15.7	476	1	US-08-313-075A-30	Sequence 30, Appl
33	52	15.7	553	4	US-08-997-251-2	Sequence 2, Appl
34	51.5	15.5	207	2	US-09-022-940-1	Sequence 1, Appl
35	51.5	15.5	207	3	US-09-216-386-1	Sequence 1, Appl
36	51.5	15.5	208	2	US-08-844-120-1	Sequence 1, Appl
37	51.5	15.5	208	4	US-09-213-394-1	Sequence 1, Appl
38	51.5	15.5	230	2	US-09-022-940-3	Sequence 3, Appl
39	51.5	15.5	230	3	US-09-216-386-3	Sequence 3, Appl
40	51	15.4	652	4	US-09-134-001C-3517	Sequence 3517, Ap
41	51	15.4	906	1	US-08-220-151-9	Sequence 9, Appl
42	51	15.4	906	1	US-08-413-118-9	Sequence 9, Appl
43	51	15.4	906	3	US-08-473-446-9	Sequence 9, Appl
44	50.5	15.2	118	3	US-08-965-904B-2	Sequence 2, Appl
45	50.5	15.2	118	4	US-08-934-131-1	Sequence 1, Appl

#### ALIGNMENTS

RESULT 1  
US-08-881-706-2  
; Sequence 2, Application US/08881706  
; Patent No. 6245969  
; GENERAL INFORMATION:  
; APPLICANT: Chong, Joane  
; APPLICANT: Li, Jianming  
; TITLE OF INVENTION: Receptor Kinase BIN1  
; FILE REFERENCE: 07251/022001  
; CURRENT APPLICATION NUMBER: US/08/881,706  
; CURRENT FILING DATE: 1997-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 1196  
; TYPE: PRT  
; ORGANISM: Arabidopsis sp.  
US-08-881-706-2

Query Match 100.0%; Score 332; DB 4; Length 1196;  
Best Local Similarity 100.0%; Pred. No. 3.6e-38;  
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AANFIAGKRYVYIKNDGKMKCHGAGNLFPGIRSEQNLRLSTRNPNITSRVYGHTS 60  
DB 588 AANFIAGKRYVYIKNDGKMKCHGAGNLFPGIRSEQNLRLSTRNPNITSRVYGHTS 647  
QY 61 PT 62  
DB 648 PT 649

RESULT 2  
US-08-568-459A-12  
; Sequence 12, Application US/08568459A  
; Patent No. 5849306  
; GENERAL INFORMATION:  
; APPLICANT: Sim, Kim L.  
; APPLICANT: Chitnis, Chetan  
; APPLICANT: Miller, Louis H.  
; APPLICANT: Peterson, David S.  
; APPLICANT: Su, Xin-zhaun  
; APPLICANT: Wellens, Thomas E.  
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Knobbe Martens Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach

STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/568,459A  
FILING DATE: 07-DEC-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelisen, Ned  
REGISTRATION NUMBER: 29,655  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2710 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium falciparum  
US-08-568-459A-12

Query Match 18.1%; Score 60; DB 2; Length 2710;  
Best Local Similarity 31.4%; Pred. No. 32;  
Matches 16; Conservative 6; Mismatches 15; Indels 14; Gaps 3;

QY 16 DGMKECHGAGNLEFQGISRSEQLRSTRNPC-----NITSRYVGGH 58  
Db 2280 DLENAACRGAG--IFSGIRKDE--WKCRNVCGYVCKPENVNGEAKGKH 2324

RESULT 3  
US-08-487-826B-12  
Sequence 12, Application US/08487826B  
GENERAL INFORMATION:  
PATENT No. 5993827  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelisen, Ned  
REGISTRATION NUMBER: 29,655

REFERENCE/DOCKET NUMBER: NIH121.001CPI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2710 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium falciparum  
US-08-487-826B-12

Query Match 18.1%; Score 60; DB 2; Length 2710;  
Best Local Similarity 31.4%; Pred. No. 32;  
Matches 16; Conservative 6; Mismatches 15; Indels 14; Gaps 3;

QY 16 DGMKECHGAGNLEFQGISRSEQLRSTRNPC-----NITSRYVGGH 58  
Db 2280 DLENAACRGAG--IFSGIRKDE--WKCRNVCGYVCKPENVNGEAKGKH 2324

RESULT 4  
US-09-210-288-12  
Sequence 12, Application US/09210288  
GENERAL INFORMATION:  
PATENT No. 6392026  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/210,288  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Fuller, Michael  
REGISTRATION NUMBER: 36,516  
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2710 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium falciparum  
US-09-210-288-12

Query Match 18.1%; Score 60; DB 4; Length 2710;  
Best Local Similarity 31.4%; Pred. No. 32;  
Matches 16; Conservative 6; Mismatches 15; Indels 14; Gaps 3;  
QY 16 DGMKECHGAGNLEFGIRSEQLNRLSTRNPC-----NITSRYVGGH 58  
DB 2280 DGLMNAICRGAG---IFSGIRKDE---WKCRNVCGYVCKPEVNVGEAKGKH 2324

RESULT 5  
US-08-487-826B-14  
Sequence 14, Application US/08487826B  
Patent No. 5993827  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487, 826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelisen, Ned  
REGISTRATION NUMBER: 29, 655  
REFERENCE/DOCKET NUMBER: NIH121.001CP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3060 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-487-826B-14  
Query Match 18.1%; Score 60; DB 2; Length 3060;  
Best Local Similarity 31.4%; Pred. No. 37;  
Matches 16; Conservative 6; Mismatches 15; Indels 14; Gaps 3;  
QY 16 DGMKECHGAGNLEFGIRSEQLNRLSTRNPC-----NITSRYVGGH 58  
DB 2278 DGLMNAICRGAG---IFSGIRKDE---WKCRNVCGYVCKPEVNVGEAKGKH 2322

RESULT 6  
US-08-244-646-15  
Sequence 15, Application US/08244646  
Patent No. 5744692  
GENERAL INFORMATION:  
APPLICANT: Cervone, Felice  
APPLICANT: De Lorenzo, Giulia  
APPLICANT: Salvi, Giovanni

APPLICANT: Albersheim, Peter  
APPLICANT: Davyall, Alan  
APPLICANT: Bergmann, Carl  
TITLE OF INVENTION: Nucleotide Sequences Coding An  
TITLE OF INVENTION: Endopolygalacturonase Inhibitor  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sally A. Sullivan  
STREET: 5370 Mannheim Circle Suite 201  
CITY: Boulder  
STATE: CO  
COUNTRY: US  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/244, 646  
FILING DATE: 06-JUN-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IT RM 91A 000915  
FILING DATE: 06-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/IT/00158  
FILING DATE: 04-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Sullivan, Sally A.  
REGISTRATION NUMBER: 32, 064  
REFERENCE/DOCKET NUMBER: 19-94  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 342 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-244-646-15  
Query Match 17.5%; Score 58; DB 1; Length 342;  
Best Local Similarity 34.5%; Pred. No. 43;  
Matches 20; Conservative 7; Mismatches 21; Indels 10; Gaps 3;  
QY 3 NFIAKRYVYIKDKMKECHGAGNLEFG---GIRSEQLNRLSTRNPCNITSRYVGG 56  
DB 236 NMLEGASVLFSGSDKNTKTKIHLAKNSIAFDLGRVGL-SKVLNGLDLRN-----NRIYV 287

RESULT 7  
US-08-592-936B-21  
Sequence 21, Application US/08592936B  
Patent No. 5783393  
GENERAL INFORMATION:  
APPLICANT: Kellogg, Jill A.  
APPLICANT: Bestwick, Richard R.  
TITLE OF INVENTION: PLANT TISSUE / STAGE SPECIFIC PROMOTERS FOR  
TITLE OF INVENTION: REGULATED EXPRESSION OF TRANSGENES IN PLANTS  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/592,936B  
FILING DATE: 29-JAN-1996  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Evans, Susan T.  
REGISTRATION NUMBER: 38,443  
REFERENCE/DOCKET NUMBER: 4257-0012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 324-0880  
TELEFAX: (650) 324-0960  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 342 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: predicted amino acid coding sequence  
INDIVIDUAL ISOLATE: of SEQ ID NO:20  
US-08-592-936B-21

Query Match 17.5%; Score 58; DB 1; Length 342;  
Best Local Similarity 34.5%; Pred. No. 4.3; Indels 10; Gaps 3;  
Matches 20; Conservative 7; Mismatches 21

3 NFIAGKRYVYIINDGMKKECHGAGNLEFQ---GIRSEQLNRLSTRNPNCTSRVYG 56  
Db 236 NMLEGDASTVLPESDKNTKIKHAKSLAFDGLKVL-SKINGLDLRN-----NRIYG 287

RESULT 8  
US-09-111-573-21  
Sequence 21, Application US/09111573  
Patent No. 5929302  
GENERAL INFORMATION:  
APPLICANT: Kellogg, Jill A.  
APPLICANT: Bestwick, Richard K.  
TITLE OF INVENTION: PLANT TISSUE / STAGE SPECIFIC PROMOTERS FOR  
TITLE OF INVENTION: REGULATED EXPRESSION OF TRANSGENES IN PLANTS  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Denlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/111,573  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/592,936  
FILING DATE: 29-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Evans, Susan T.  
REGISTRATION NUMBER: 38,443  
REFERENCE/DOCKET NUMBER: 4257-0012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 324-0880  
TELEFAX: (650) 324-0960  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 342 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: predicted amino acid coding sequence  
INDIVIDUAL ISOLATE: of SEQ ID NO:20  
US-09-111-573-21

Query Match 17.5%; Score 58; DB 2; Length 342;  
Best Local Similarity 34.5%; Pred. No. 4.3; Indels 10; Gaps 3;  
Matches 20; Conservative 7; Mismatches 21

3 NFIAGKRYVYIINDGMKKECHGAGNLEFQ---GIRSEQLNRLSTRNPNCTSRVYG 56  
Db 236 NMLEGDASTVLPESDKNTKIKHAKSLAFDGLKVL-SKINGLDLRN-----NRIYG 287

RESULT 9  
US-09-220-528-64  
Sequence 64, Application US/09220528A  
Patent No. 6284540  
GENERAL INFORMATION:  
APPLICANT: Milbrandt, Jeffrey D.  
APPLICANT: Balch, Robert H.  
TITLE OF INVENTION: Artemin, A No. 6284540el Neurotrophic Factor  
FILE REFERENCE: 6029-7998  
CURRENT APPLICATION NUMBER: US/09/220,528A  
CURRENT FILING DATE: 1998-12-24  
EARLIER APPLICATION NUMBER: 09/218,698  
EARLIER FILING DATE: 1998-12-22  
EARLIER APPLICATION NUMBER: 60/108,148  
EARLIER FILING DATE: 1998-11-12  
EARLIER APPLICATION NUMBER: 09/163,283  
EARLIER FILING DATE: 1998-09-29  
NUMBER OF SEQ ID NOS: 120  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 64  
LENGTH: 397  
TYPE: PRT  
ORGANISM: Murine  
US-09-220-528-64

Query Match 17.5%; Score 58; DB 4; Length 397;  
Best Local Similarity 44.4%; Pred. No. 5.2; Indels 4; Gaps 1;  
Matches 12; Conservative 4; Mismatches 7

22 CHGAGNLEFQGRISQNLRLSTRNPC 48  
Db 315 CRGSGNLD---ECQELERSFQNPC 337

RESULT 10  
US-08-533-669A-18  
Sequence 18, Application US/08533669A  
Patent No. 5834592  
GENERAL INFORMATION:  
APPLICANT: Corixa Corporation  
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/533,669A  
FILING DATE: 22-SEP-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.420  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 732 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-533-669A-18

Query Match 17.3%; Score 57.5; DB 2; Length 732;  
Local Similarity 37.0%; Pred. No. 13;  
Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;

QY 14 KNDGKKECHGAGNLEFGGIRSEQLNR--LSTRNPNCTISRYVG 56  
DB 269 KKGDKKK---KKIKIKYIDKEINKTKPIWTRNPDITNEBYG 310

RESULT 11  
US-09-307-143-4  
Sequence 4, Application US/09307143  
Patent No. 6335157  
GENERAL INFORMATION:  
APPLICANT: Gonzalez C.  
APPLICANT: Lange, B.  
TITLE OF INVENTION: METHODS BASED ON LOCALIZATION OF HSP90 TO THE  
FILE REFERENCE: 9882-003  
CURRENT APPLICATION NUMBER: US/09/307,143  
CURRENT FILING DATE: 1999-05-07  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 4  
LENGTH: 732  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-307-143-4

Query Match 17.3%; Score 57.5; DB 4; Length 732;  
Local Similarity 37.0%; Pred. No. 13;  
Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;

QY 14 KNDGKKECHGAGNLEFGGIRSEQLNR--LSTRNPNCTISRYVG 56  
DB 269 KKGDKKK---KKIKIKYIDKEINKTKPIWTRNPDITNEBYG 310

RESULT 12  
US-09-183-861-18  
Sequence 18, Application US/09183861  
Patent No. 6365165  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Webb, John R.  
APPLICANT: Dillon, David C.  
APPLICANT: Skeiky, Yasir A.W.  
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USEIN THE THERAPY AND  
NUMBER OF SEQUENCES: 87  
CORRESPONDENCE ADDRESS:  
ADDRESSER: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington

COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/183,861  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/022,765  
FILING DATE: 12-FEB-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.420C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 732 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-09-183-861-18

Query Match 17.3%; Score 57.5; DB 4; Length 732;  
Local Similarity 37.0%; Pred. No. 13;  
Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;

QY 14 KNDGKKECHGAGNLEFGGIRSEQLNR--LSTRNPNCTISRYVG 56  
DB 269 KKGDKKK---KKIKIKYIDKEINKTKPIWTRNPDITNEBYG 310

RESULT 13  
US-09-022-765-18  
Sequence 18, Application US/09022765  
Patent No. 6375955  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Webb, John R.  
APPLICANT: Dillon, David C.  
APPLICANT: Skeiky, Yasir A.W.  
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USEIN THE THERAPY AND  
NUMBER OF SEQUENCES: 87  
CORRESPONDENCE ADDRESS:  
ADDRESSER: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/022,765  
FILING DATE: 12-FEB-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.420C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 732 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-09-022-765-18

Query Match 17.3%; Score 57.5; DB 4; Length 732;  
Best Local Similarity 37.0%; Pred. No. 13;  
Matches 17; Conservative 9; Mismatches 13; Indels 7; Gaps 2;

QY 14 KNDGKKECHGAGNLLFPGIRSEQLNR---LSTRNPNCTISRVYG 56  
Db 269 KKDQDKK---KKKIKKXIDKEILNKTPIWTRNPDITNEEYG 310

RESULT 14  
US-09-045-284A-2  
Sequence 2, Application US/09045284A  
Accession No. 6265192

GENERAL INFORMATION:

APPLICANT: Bistrup, Annette

APPLICANT: Rosen, Steven D.

APPLICANT: Hemmerich, Stefan

TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3

FILE REFERENCE: 6510-107US1

CURRENT APPLICATION NUMBER: US/09/045,284A

CURRENT FILING DATE: 1998-03-20

NUMBER OF SEQ ID NOS: 9

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 2

LENGTH: 386

TYPE: PRT

ORGANISM: Homo sapiens

US-09-045-284A-2

Query Match 17.2%; Score 57; DB 4; Length 386;  
Best Local Similarity 41.0%; Pred. No. 7;  
Matches 16; Conservative 6; Mismatches 11; Indels 6; Gaps 1;

QY 12 YIKNDGKKECHGAGNLLFPGIRSEQ-----LNRST 44  
Db 340 YEKVSRLLQKACGDAMNLLGYRHVRSSEQQRNLLDLST 378

RESULT 15

US-09-190-911-1

Sequence 1, Application US/09190911

Accession No. 6365365

GENERAL INFORMATION:

APPLICANT: Bistrup, Annette

APPLICANT: Rosen, Steven D.

APPLICANT: Tangemann, Kirsten

APPLICANT: Hemmerich, Stefan

TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3

FILE REFERENCE: 6510-107CIP

CURRENT APPLICATION NUMBER: US/09/190,911

CURRENT FILING DATE: 1998-11-12

EARLIER APPLICATION NUMBER: 09/045,284

EARLIER FILING DATE: 1998-03-20

NUMBER OF SEQ ID NOS: 8

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 1

LENGTH: 386

TYPE: PRT

ORGANISM: H. sapiens

US-09-190-911-1

QY 12 YIKNDGKKECHGAGNLLFPGIRSEQ-----LNRST 44  
Db 340 YEKVSRLLQKACGDAMNLLGYRHVRSSEQQRNLLDLST 378  
Search completed: March 10, 2003, 18:19:23  
Job time : 3.33068 secs

Query Match 17.2%; Score 57; DB 4; Length 386;  
Best Local Similarity 41.0%; Pred. No. 7;  
Matches 16; Conservative 6; Mismatches 11; Indels 6; Gaps 1;





PT plants and for modulating oocyte maturation  
PS Claim 3; Page 52; 72pp; English.

CC This is the amino acid sequence of a novel plant steroid receptor  
CC kinase, designated Bn1, which is involved in the pathway for the  
CC synthesis of the plant steroid hormone, brassinolide. 18 New  
CC Arabidopsis dwarf mutants were identified that lacked the ability  
CC to respond to brassinolide, and were named bin mutants. The bin1  
CC mutations were used to map the gene to a small interval on  
CC Arabidopsis chromosome 4. Bin1 was cloned by map-based cloning.  
CC A Bin1 polynucleotide (see AAX07356) was identified within this  
CC interval by sequencing the wild-type and mutant alleles of this  
CC nuclear acid. Overexpression of Bin1 in transgenic plants provides  
CC plants characterized as having enhanced disease resistance,  
CC increased plant yield or vegetative biomass and increased seed  
CC yield. Bin1 expression may also increase resistance to pesticides.  
CC Inhibition of Bin1, e.g. using antisense oligonucleotides (AON), is  
CC used to render plants male-sterile, and to reduce their stature or  
CC yield, e.g. for creating dwarf varieties. Since Bin1 homologues  
CC may be involved in regulation of the menstrual cycle and uterine  
CC function, Bin1, antibodies and AON may be useful as contraceptives,  
CC for improving success of in vitro fertilisation and to prevent  
CC premature labour. Transgenic animals are also provided, and are  
CC models for studying steroid-receptor interactions or can be used  
CC to screen for therapeutic agents.

XX Sequence 1196 AA;

SO Query Match 100.0%; Score 332; DB 20; Length 1196;  
Best Local Similarity 100.0%; Pred. No. 2.2e-37;  
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AANFAGKRYVYIKNDGKKECHGAGNLEFGIRSEQLNRLSTNPNCTISRYGGHTS 60  
DB 588 AANFAGKRYVYIKNDGKKECHGAGNLEFGIRSEQLNRLSTNPNCTISRYGGHTS 647

QY 61 PT 62  
DB 648 PT 649

RESULT 2  
AAB67443  
ID AAB67443 standard; Protein; 1196 AA.

AC AAB67443;

XX 15-MAY-2001 (first entry)

DE Amino acid sequence of an Arabidopsis Br1 protein.

XX Disease resistance protein; Xa21; RKK gene; transgenic plant;  
KW Xanthomonas; plant pathogen; Br1 protein; RCH10 protein.

XX Arabidopsis sp.

XX WO200109283-A2.

XX 08-FEB-2001.

XX 28-JUL-2000; 2000WO-US20714.

XX 28-JUL-1999; 99US-036313.

XX (REGC ) UNIV CALIFORNIA.

XX (SALK ) SALK INST BIOLOGICAL STUDIES.

XX Ronald P, He Z, Chory J, Lamb C, Li J;

XX WPI; 2001-159858/16.

XX N-PSDB; AAF54982.

PT Chimeric plant receptors comprising a polynucleotide encoding a RKK  
PT receptor containing a heterologous extracellular domain and a kinase  
PT domain from a Xa21 polypeptide, useful for modulating plant responses  
PT to pathogens

PS Disclosure; Page 38-42; 47pp; English.

CC The present sequence represents a Br1 protein. The specification  
CC describes chimeric receptors which are used for modulating plant  
CC responses to pathogens. The receptors comprise a heterologous  
CC extracellular domain (e.g. from a Br1 protein or RCH10 protein) and  
CC a kinase domain (e.g. from disease resistance protein Xa21). The Xa21  
CC gene is a member of disease resistance genes referred to as RKK genes.  
CC When Xa21 is present in a transgenic plant, it confers resistance to  
CC Xanthomonas spp.. The plant receptors containing heterologous domains  
CC are useful for modulating plant responses to pathogens including viruses,  
CC bacteria, nematodes, fungi or insects. The nucleic acids can be used to  
CC confer desired traits on essentially any plant.

XX Sequence 1196 AA;

SO Query Match 100.0%; Score 332; DB 22; Length 1196;  
Best Local Similarity 100.0%; Pred. No. 2.2e-37;  
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AANFAGKRYVYIKNDGKKECHGAGNLEFGIRSEQLNRLSTNPNCTISRYGGHTS 60  
DB 588 AANFAGKRYVYIKNDGKKECHGAGNLEFGIRSEQLNRLSTNPNCTISRYGGHTS 647

QY 61 PT 62  
DB 648 PT 649

RESULT 3  
AAE19490  
ID AAE19490 standard; Protein; 1196 AA.

XX AAE19490;

XX 31-MAY-2002 (first entry)

DE Brassinosteroid receptor protein encoded by Br1 DNA.

XX Plant receptor protein; disease resistance protein; plant development;

KW RKK protein; disease resistance; cell free assay; gene therapy;

KW Br1 DNA; Brassinosteroid protein.

XX Unidentified.

XX Key Location/Qualifiers

XX MISC-difference 162 /note= "Encoded by TTT"

XX WO200210367-A1.

XX 07-FEB-2002.

XX 28-JUL-2000; 2000WO-US20604.

XX 28-JUL-2000; 2000WO-US20604.

XX (SALK ) SALK INST BIOLOGICAL STUDIES.

XX Chory J, Lamb C, He Z;

XX WPI; 2002-227084/28.

XX N-PSDB; AAD30800.

XX Heterologous polynucleotide encoding chimeric plant receptors for  
PT controlling plant development and disease resistance, has leucine-rich  
PT repeat domain, transmembrane domain, and cytoplasmic protein kinase  
PT domain

XX PS Disclosure; Page 49-50; 54pp; English.

CC The present invention relates to heterologous nucleic acid molecules

CC encoding chimeric plant receptor proteins comprising a leucine-rich

CC repeat (LRR) domain, a transmembrane domain and a kinase domain from

CC cytoplasmic RIK (disease resistance proteins) proteins such as Xa21.

CC The nucleic acid sequences are useful for identifying ligands for

CC receptor or receptor-like kinase. The chimeric receptors are useful

CC for controlling plant development and/or disease resistance. They are

CC used in cell free assay useful for determining the ability of a test

CC compound to bind to or modulate the activity or expression of the

CC receptor. Sequences of the invention are also used in gene therapy.

CC The present sequence is brassinosteroid receptor protein encoded by

CC btl DNA. This sequence is used in the invention.

XX SQ Sequence 1196 AA;

Query Match 100.0%; Score 332; DB 23; Length 1196;

Best Local Similarity 100.0%; Pred. No. 2.2e-37; Mismatches 0; Gaps 0;

Phes 62; Conservative 0; Indels 0; Gaps 0;

QY 1 AANFIAGKRYVYIKNDGKKECHGAGNILEFGIRSEQLRLSTRNPNCTITSRYGHTS 60

DB 588 AANFIAGKRYVYIKNDGKKECHGAGNILEFGIRSEQLRLSTRNPNCTITSRYGHTS 647

QY 61 PT 62

DB 648 PT 649

RESULT 4

ABP00429

ID ABP00429 standard; Protein; 93 AA.

AC ABP00429;

DT 24-JUN-2002 (first entry)

XX Human ORFX protein sequence SEQ ID NO:840.

XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;

XX hyperproliferative disorder; psoriasis; benign tumor; hemorhage;

XX degenerative disorder; osteoarthritis; neurodegenerative disorder;

XX cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;

XX hypertension; hypothyroidism; cholesterol ester storage disease;

XX immune deficiency; immune disorder; infectious disease;

XX autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;

XX myasthenia gravis.

XX Homo sapiens.

OS WO200192523-A2.

PN 06-DEC-2001.

XX 29-MAY-2001; 2001WO-US10836.

XX 30-MAY-2000; 2000US-206132P.

XX 29-AUG-2000; 2000US-228716P.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach MD;

XX WPI; 2002-106308/14.

XX N-PSDB; ABL16181.

XX Novel human polypeptides and polynucleotides useful for diagnosing,

XX preventing and treating cardiovascular disease, neurodegenerative,

XX hyperproliferative disorders and autoimmune disorders

XX Disclosure; SEQ ID 840; 1037pp; English.

XX CC The present invention describes substantially purified human proteins

CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1

CC in the specification). ABLN15762 to ABLN27252 encode the human ORFX

CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for

CC treating or preventing a pathology associated with an ORFX-associated

CC disorder in humans, and in the manufacture of a medicament for treating a

CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide

CC sequences can be used in gene therapy. ORFX sequences can be used in the

CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,

CC psoriasis, benign tumors, keloid, degenerative disorders, haemorrhage,

CC osteoarthritis, neurodegenerative disorders, disorders related to organ

CC transplantation, cardiovascular diseases, diabetes mellitus, systemic

CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester

CC storage disease, various immune deficiencies and disorders, infectious

CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid

CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host

CC disease and autoimmune inflammatory eye disease. ORFX proteins are also

CC useful for treating burns, incisions, ulcers, for treating osteoporosis,

CC bone degenerative disorders, or periodontal disease, and for gut

CC protection or regeneration and treatment of lung or liver fibrosis,

CC reperfusion injury in various tissues and conditions resulting from

CC systemic cytokine damage.

CC N.B. The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 93 AA;

Query Match 43.8%; Score 145.5; DB 23; Length 93;

Best Local Similarity 51.0%; Pred. No. 1.6e-12;

Matches 26; Conservative 11; Mismatches 13; Indels 1; Gaps 1;

QY 5 IAGKRYVYIKNDGKKECHGAGNILEFGIRSEQLRLSTRNPNCTITSRY 55

DB 44 VIGRPVYLRNDELBSBCHGKGLIFETSRPBLSPMPKELCNFT-RVY 93

RESULT 5

AAAG79244

ID AAAG79244 standard; Protein; 1121 AA.

AC AAAG79244;

DT 03-JAN-2002 (first entry)

XX Amino acid sequence of a gene concerning brassinosteroid sensitivity.

XX OseBRI1; brassinosteroid sensitivity; d61 locus; rice;

XX internode elongation; internode cell; lamina joint.

XX Oryza sativa.

OS WO200173036-A1.

PN 04-OCT-2001.

XX 30-MAR-2001; 2001WO-JP02770.

XX 31-MAR-2000; 2000JP-0101276.

XX (NAAG-) NAT INST AGROBIOLOGICAL SCI.

XX Tanaka H, Kayano T, Matsunaka M;

XX WPI; 2001-616505/71.

XX N-PSDB; AAI65842.

XX Gene relating to brassinosteroid-sensitivity of plants, useful in

XX controlling growth and development of transformants including rice to

XX improve harvest and crop yield for animal feed or dwarfism to enhance

XX ornamental effect

PS Claim 1; Page 72-80; 87pp; Japanese.

CC The present sequence is encoded by a gene, designated OsBRI1, which  
CC enhances the brassinosteroid sensitivity of plants. The OsBRI1 gene is  
CC located to the d61 locus. The gene is involved in the growth and  
CC development of rice, for example, internode elongation via the induction  
CC of the elongation of internode cells and bending of lamina joints.  
CC The OsBRI1 gene is useful in controlling growth and development of  
CC transformants. Transformant rice plants can be obtained to improve  
CC harvest and crop yield for providing more animal feed in agriculture,  
CC or plants with dwarfism can be produced to enhance ornamental effect  
CC and added value in horticulture and related industries.

CC Sequence 1121 AA;

Query Match 42.3%; Score 140.5; DB 22; Length 1121;  
Best Local Similarity 46.6%; Pred. No. 1.9e-10;  
Matches 27; Conservative 11; Mismatches 19; Indels 1; Gaps 1;

5 IAGKRYVYIKNDGKKECHGAGNLEFGIRSEQLNRLSTRNPNTTSRYGGHTSPT 62  
ID IAGKRYVYIKNDGKKECHGAGNLEFGIRSEQLNRLSTRNPNTTSRYGGHTSPT 62  
520 IAGKRYVYIKNDGKKECHGAGNLEFGIRSEQLNRLSTRNPNTTSRYGGHTSPT 576

RESULT 6

ABB91366 standard; Protein; 1166 AA.

ABB91366;

31-MAY-2002 (first entry)

Herbicideally active polypeptide SEQ ID NO 577.

Herbicideal; plant; agriculture; herbicide.

Arabidopsis thaliana.

WO200210210-A2.

07-FEB-2002.

28-AUG-2001; 2001WO-EP09892.

28-AUG-2001; 2001WO-EP09892.

(FARB ) BAYER AG.

Tietjen K, Weidler M;

WPI; 2002-269010/31.

Identifying plant target proteins for herbicideally active compounds,  
comprising aligning and comparing nucleic acid or amino acid sequences  
from plant with nucleic acid or amino acid sequences from non-plant  
organisms -

Claim 5; SEQ ID NO 577; 261pp + Sequence Listing; English.

The invention relates to identifying target proteins  
comprising aligning and comparing nucleic acid or amino acid sequences from plant  
with nucleic acid or amino acid sequences from non-plant organisms using  
suitable search parameters, where plant sequences having an E-value  
greater by a factor of 3 than the E-value of most similar non-plant  
sequences are selected. The polypeptides or nucleic acids encoding them  
are useful for identifying modulators. The identified modulators are  
useful as herbicides.

Sequence 1166 AA;

Query Match 38.6%; Score 128; DB 23; Length 1166;  
Best Local Similarity 43.1%; Pred. No. 1.1e-08;

Matches 25; Conservative 15; Mismatches 16; Indels 2; Gaps 2;

5 IAGKRYVYIKNDGKKECHGAGNLEFGIRSEQLNRLSTRNPNTTSRYGGHTSPT 62  
ID IAGKRYVYIKNDGKKECHGAGNLEFGIRSEQLNRLSTRNPNTTSRYGGHTSPT 62  
579 VSGKQFAFVNEG-GTDCRAGAGLVEFGIRARLERLPMVHSCPAT-RITYSGMTWT 634

RESULT 7

ABB92294 standard; Protein; 1164 AA.

ABB92294;

31-MAY-2002 (first entry)

Herbicideally active polypeptide SEQ ID NO 1505.

Herbicideal; plant; agriculture; herbicide.

Arabidopsis thaliana.

WO200210210-A2.

07-FEB-2002.

28-AUG-2001; 2001WO-EP09892.

28-AUG-2001; 2001WO-EP09892.

(FARB ) BAYER AG.

Tietjen K, Weidler M;

WPI; 2002-269010/31.

Identifying plant target proteins for herbicideally active compounds,  
comprising aligning and comparing nucleic acid or amino acid sequences  
from plant with nucleic acid or amino acid sequences from non-plant  
organisms -

Claim 5; SEQ ID NO 1505; 261pp + Sequence Listing; English.

The invention relates to identifying target proteins  
comprising aligning and comparing nucleic acid or amino acid sequences from plant  
with nucleic acid or amino acid sequences from non-plant organisms using  
suitable search parameters, where plant sequences having an E-value  
greater by a factor of 3 than the E-value of most similar non-plant  
sequences are selected. The polypeptides or nucleic acids encoding them  
are useful for identifying modulators. The identified modulators are  
useful as herbicides.

Sequence 1164 AA;

Query Match 34.9%; Score 116; DB 23; Length 1164;  
Best Local Similarity 40.0%; Pred. No. 5.6e-07;  
Matches 22; Conservative 15; Mismatches 16; Indels 2; Gaps 2;

5 IAGKRYVYIKNDGKKECHGAGNLEFGIRSEQLNRLSTRNPNTTSRYGGHT 59  
ID IAGKRYVYIKNDGKKECHGAGNLEFGIRSEQLNRLSTRNPNTTSRYGGHT 59  
579 VSGKQFAFVNEG-GTDCRAGAGLVEFGIRARLERLPMVHSCPAT-RITYSGMT 631

RESULT 8

ABB25323 standard; Protein; 919 AA.

ABB25323;

27-NOV-2000 (first entry)

Eucalyptus grandis cell signalling involved protein SEQ ID NO:642.

KW	Eucalyptus grandis; Pinus radiata; Monterey pine; plant; modification;
KM	plant cell signalling; modulation; transgenic plant; pathogen; growth;
KM	environmental change; development; cell proliferation; differentiation;
KX	elongation; survival; disease resistance; nutrient metabolism.
XX	
OS	Eucalyptus grandis.
PN	MO200042171-A1.
PD	20-JUL-2000.
PX	
PF	11-JAN-2000; 2000WO-US00724.
XX	
PR	12-JAN-1999; 99US-0228986.
XX	
PR	01-NOV-1999; 99US-0162866.
PA	(GENE-) GENESIS RES & DEV CORP LTD.
PI	Strabala TJ, Nieuwenhuizen NJ;
DY	PI; 2000-476052/41.
PT	Isolated polynucleotide encoding a polypeptide involved in cell
PT	signaling used for generating transgenic plants with modified responses
PT	to external signals -
PS	Claim 3; Page 291-293; 527pp; English.
CC	
CC	AAA79263 to AAA79736 and AAB25100 to AAB25570 represent polynucleotide
CC	and protein sequences isolated from Eucalyptus (Eucalyptus grandis) or
CC	pine (Pinus radiata also known as Monterey pine). The protein sequences
CC	are involved in cell signalling. The polynucleotide and protein
CC	sequences can be used to modify the response of plant cells to external
CC	signals e.g. environmental changes or pathogens during the growth and
CC	development of a plant. They can be used to modify cell proliferation,
CC	differentiation, elongation and survival, resistance to disease and
CC	nutrient metabolism. Examples of modifications which can be produced are
CC	altered fruit ripening and senescence of leaves and flowers e.g. to
CC	delay senescence and prolong the life of cut flowers or enhance
CC	senescence of reproductive organs to engineer sterile plants. Other
CC	modifications can be used to delay senescence in selected cell types or
CC	organs providing fruit and vegetables which have a longer shelf life
CC	between harvest and consumption, or to decrease branching frequency in
CC	forest tree species giving long stretches of valuable knot-free clear
CC	wood which can be used in solid timber furniture and veneers.
XX	
SQ	Sequence 919 AA:
TY	Match 29.4%; Score 97.5; DB 21; Length 919;
CC	Local Similarity 39.3%; Pred. No. 0.00017;
Matches	22; Conservative 10; Mismatches 21; Indels 3; Gaps 2;
OY	2 ANFIAGKRYVYITKDGMKKCHGAGNLTLEFFOGRISPOLNRILSTRNPCNTSRVGG 57
DB	329 AGIPSGNTLVFRNVG--NTCKGVGLLEFAGIRPERLLQVSLRCP-AMRTS 361
ID	AAB25490 standard; Protein; 1133 AA.
AC	
AA	AAB25490;
DT	27-NOV-2000 (first entry)
DE	Eucalyptus grandis cell signalling involved protein SEQ ID NO:809.
XX	
KW	Eucalyptus grandis; Pinus radiata; Monterey pine; plant; modification;
KM	plant cell signalling; modulation; transgenic plant; pathogen; growth;
KM	environmental change; development; cell proliferation; differentiation;
KX	elongation; survival; disease resistance; nutrient metabolism.
XX	
OS	Bucalyptus grandis.

PX	NN	WO200042171-A1.
PX	PD	20-JUL-2000.
PX	PF	11-JAN-2000; 200OWO-US00724.
PX	PR	12-JAN-1999; 99US-0228986.
PX	PP	01-NOV-1999; 99US-0162866.
PA	(GENE-)	GENESIS RES & DEV CORP LTD.
XX	Strabala TJ,	Nieuwenhuizen NJ;
XX	PI	WPI; 2000-476052/41.
XX	DR	
PT	Isolated polynucleotide encoding a polypeptide involved in cell	
PT	signaling used for generating transgenic plants with modified responses	
PT	to external signals -	
PS	Claim 3; Page 378-380; 527pp; English.	
PX	AAA79263 to AAA79736 and AAB25100 to AAB25570 represent polynucleotide	
CC	and protein sequences isolated from eucalyptus (Eucalyptus grandis) or	
CC	pine (Pinus radiata also known as Monterey pine). The protein sequences	
CC	are involved in cell signalling. The polynucleotide and protein	
CC	sequences can be used to modify the response of plant cells to external	
CC	signals e.g. environmental changes or pathogens during the growth and	
CC	development of a plant. They can be used to modify cell proliferation,	
CC	differentiation, elongation and survival, resistance to disease and	
CC	nutrient metabolism. Examples of modifications which can be produced are	
CC	altered fruit ripening and senescence of leaves and flowers e.g. to	
CC	delay senescence and prolong the life of cut flowers or enhance	
CC	senescence of reproductive organs to engineer sterile plants. Other	
CC	modifications can be used to delay senescence in selected cell types or	
CC	organs providing fruit and vegetables which have a longer shelf life	
CC	between harvest and consumption, or to decrease branching frequency in	
CC	forest tree species giving long stretches of valuable knot-free clear	
CC	wood which can be used in solid timber furniture and veneers.	
SQ	Sequence 1133 AA;	
Query Match	29.4%; Score 97.5; DB 21; Length 1133;	
Best Local Similarity	39.3%; Pred. No. 0.00022;	
Matches 22; Conservative 10; Mismatches 21; Indels 3; Gaps 2		
CY	2 ANFIAGKRVYIKNDGMKECHGAGNLTFFOSIRSEOLNRSTRMPCNTSRVYGG 57	
DB	543 AGIPSGNTLVFPRNVG--NTCKGVGLLFPAGIRPERLLQVPISLRTCNF-AAWYSG 595	
RESULT 10		
ID	AAB91692 standard; Protein; 1143 AA.	
XX	AAB91692;	
AC		
XX	31-MAY-2002 (first entry)	
DE	Herbicidally active polypeptide SEQ ID NO 903.	
XX		
KW	Herbicidal; plant; agriculture; herbicide.	
OS	Arabidopsis thaliana.	
XX	WO200210210-A2.	
FN		
XX	07-FEB-2002.	
PD		
XX	28-AUG-2001; 2001WO-EPO9892.	
PF		
XX	28-AUG-2001; 2001WO-EPO9892.	
DR		
XX		

[illegible][illegible]

```

RESULT 12
AAB96183
ID AAB96183 standard; Protein; 916 AA.
XX
XX AAB96183;
AC
XX
XX
DT 29-OCT-2001 (first entry)
XX
XX Putative P. abyssi protein #3.
XX
XX Hyperthermophilic archaeon; hyperthermophilic protein.
XX
XX Pyrococcus abyssi.
OS
XX
XX FR2792651-A1.
XX
XX
XX 27-OCT-2000.
XX
XX
XX 21-APR-1999; 99FR-0005034.
XX
XX
XX 21-APR-1999; 99FR-0005034.
XX
XX
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX (IFRE-) IFREMER INST FR RECH EXPL MER.
PA
XX
XX
XX Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
XX
XX Quereillon J, Weissenbach J, Saurin W, Hellig R;
XX
XX
XX WPI; 2001-126236/14.
XX
XX
XX New nucleotide sequences isolated from Pyrococcus abyssi encode
XX
XX proteins useful in industry -
XX
XX
XX Claim 7; Pages 820-823; 1657BP; French.
XX
XX
XX The present invention relates to the genomic sequence of Pyrococcus
XX
XX abyssi (see AA086431 and AA041223-7) and P. abyssi proteins. P. abyssi is

```

CC a hyperthermophilic archaeon, which is isolated from deep-sea  
 CC hydrothermal vents. The present sequence is one such P. abyssi protein.  
 CC The proteins of the present invention have various potential industrial  
 CC uses, since the proteins are stable at very high temperatures, some up to  
 CC 110 degrees centigrade.  
 CC Note: This patent is in the same patent family as WO2000065062, which  
 CC contains additional sequences as shown in AAB9132-AAB9143,  
 CC AAH75903-AAH75920 and AAG66436.  
 XX  
 SQ Sequence 916 AA;  
 Query Match 19.0%; Score 63; DB 22; Length 916;  
 Best Local Similarity 39.1%; Pred. No. 12;  
 Matches 18; Conservative 5; Mismatches 9; Indels 14; Gaps 3;  
 QY 10 YVYIKNDGKKE-----CHGAGNLL-----EPQG--IRSEQLNR 41  
 DB 821 YVLAGTEGAMKERTPGTCHGAGVLSRKAAATRGVGRIRIGELLNR 866  
 F 13  
 A 580  
 ID AAE12580 standard; Protein; 466 AA.  
 XX  
 AC AAE12580;  
 XX  
 DT 03-JAN-2002 (first entry)  
 XX  
 DE Pseudomonas syringae pv. tomato (Pto) DC3000 EEL ORF1 encoded protein.  
 XX  
 KW Conserved Effector Loci; CEL; cytosolic; antibacterial; gene therapy;  
 KW Exchangeable Effector Loci; EEL; disease resistance; transgenic plant;  
 KW eukaryotic cell death; cancer.  
 XX  
 OS Pseudomonas syringae.  
 XX  
 PN WO200175066-A2.  
 PD 11-OCT-2001.  
 XX  
 PF 03-APR-2001; 2001WO-US10698.  
 XX  
 PR 03-APR-2000; 2000US-194160P.  
 PR 11-AUG-2000; 2000US-224604P.  
 PR 17-NOV-2000; 2000US-249548P.  
 XX  
 PA (CORR ) CORNELL RES FOUND INC.  
 PA (UINE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.  
 PA (UINE-) UNIV NEBRASKA.  
 XX  
 PL Colimer A, Alfano JR, Charkowski AO,  
 DR MPI: 2001-639361/73.  
 DR N-PSDB; AAD20415.  
 XX  
 PT New nucleic acid molecules encoding proteins or polypeptides of  
 PT Pseudomonas Conserved Effector Loci and Exchangeable Effector Loci  
 PT genomic sequences, for imparting disease resistance to plants  
 XX  
 PS Claim 8; Page 32-33; 217pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid molecule comprising a  
 CC nucleotide sequence encoding proteins or polypeptides of Pseudomonas  
 CC Conserved Effector Loci (CEL) and Exchangeable Effector Loci (EEL)  
 CC genomic sequences. CEL and EEL DNA are useful for imparting disease  
 CC resistance to a plant, by transforming a plant cell with the nucleic acid  
 CC and regenerating a transgenic plant from the transformed plant cell,  
 CC where the transgenic plant expresses a heterologous DNA molecule under  
 CC conditions effective to impart disease resistance, or by treating a plant  
 CC with an isolated protein or polypeptide, by applying the protein or  
 CC polypeptide in an isolated form or by applying a non-pathogenic bacteria  
 CC which secretes the protein or polypeptide, under conditions effective to  
 CC impart disease resistance to the treated plant. CEL and EEL proteins

CC are useful for causing eukaryotic cell death, by introducing a cytotoxic  
 CC Pseudomonas protein into a eukaryotic cell under conditions effective to  
 CC cause cell death. CEL and EEL proteins are also useful for treating a  
 CC cancerous condition, by introducing a cytotoxic Pseudomonas protein into  
 CC cancer cells of a patient under conditions effective to cause death of  
 CC cancer cells, and thus treating the cancerous condition. The method  
 CC further involves administering a targeted DNA delivery system  
 CC comprising a DNA molecule encoding the cytotoxic Pseudomonas protein,  
 CC to the patient, where the targeted DNA delivery system delivers the  
 CC DNA molecule into cancer cells and the cytotoxic Pseudomonas protein  
 CC is expressed in the cancer cells. The present sequence is  
 CC Pseudomonas syringae pv. tomato (Pto) DC3000 EEL ORF1 protein.  
 XX  
 SQ Sequence 466 AA;  
 Query Match 18.2%; Score 60.5; DB 22; Length 466;  
 Best Local Similarity 26.4%; Pred. No. 11;  
 Matches 19; Conservative 11; Mismatches 31; Indels 11; Gaps 2;  
 QY 1 AANFIAGKRYVYIKNDG-----KKECHGAGNLLFPGTISEQLN--RLETRPCN 49  
 DB 387 AMNYVAAEKIRLSKPEKRVAFVGAHTATSCDGVGLAEHGVSLVIDGLKSRATVD 446  
 QY 50 ITSRYVGGTSP 61  
 DB 447 INVKNYGGKLN 458  
 RESULT 14  
 ABB63435  
 ID ABB63435 standard; Protein; 725 AA.  
 XX  
 AC ABB63435;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 17097.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PT Venter JC, Adams M, Li PWD, Myers EW,  
 DR MPI: 2001-656860/75.  
 DR N-PSDB; ABL07538.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions  
 XX  
 PS Disclosure; SEQ ID NO 17097; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABJ16176-ABJ16511), expressed DNA  
 CC sequences (ABJ01840-ABJ16175) and the encoded proteins  
 CC (ABB7757-ABB72072).

CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
 SQ Sequence 725 AA;

Query Match 18.2%; Score 60.5; DB 22; Length 725;  
 Best Local Similarity 38.0%; Pred. No. 20;  
 Matches 19; Conservative 5; Mismatches 23; Indels 3; Gaps 1;

QY 2 ANFIAGKRYIYIKDGMKECHGAGNLEFGIRSEQLNRLSTRNPNIT 51  
 Db 642 ARFKQGRDYRYEYDNGINP--GAFNPLQQLRIKVTARLTCDSNDRLT 688

RESULT 15

ABB71725  
 ID ABB71725 standard; Protein; 1341 AA.

XX ABB71725;

DI 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 41967.

XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

XX 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL15828.

XX

PS Disclosure; SEQ ID NO 41967; 21pp + Sequence listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1341 AA;

Query Match 18.2%; Score 60.5; DB 22; Length 1341;

Best Local Similarity 40.5%; Pred. No. 44; Mismatches 12; Indels 9; Gaps 1;

Matches 17; Conservative 4; Mismatches 12; Indels 9; Gaps 1;

QY 7 GKRYYIKNDGMKECHGAGNLEFGIRSEQLNRLSTRNPNIT 48  
 Db 676 GKRFRYIENTIMKLPQHGWNLL-----LHRLSYPLC 708

Search completed: March 10, 2003, 18:03:34  
 Job time : 4.35135 secs